



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **A decorated tree approach to random permutations in substitution-closed classes**

Borga, Jacopo ; Bouvel, Mathilde ; Féray, Valentin ; Stufler, Benedikt

**Abstract:** We establish a novel bijective encoding that represents permutations as forests of decorated (or enriched) trees. This allows us to prove local convergence of uniform random permutations from substitution-closed classes satisfying a criticality constraint. It also enables us to reprove and strengthen permuton limits for these classes in a new way, that uses a semi-local version of Aldous' skeleton decomposition for size-constrained Galton–Watson trees.

DOI: <https://doi.org/10.1214/20-ejp469>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-191594>

Journal Article

Published Version

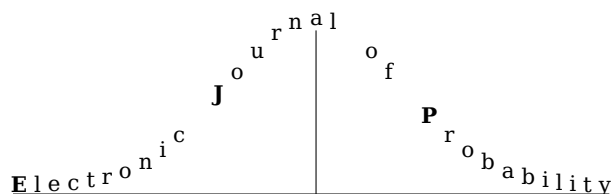


The following work is licensed under a Creative Commons: Attribution 3.0 Unported (CC BY 3.0) License.

Originally published at:

Borga, Jacopo; Bouvel, Mathilde; Féray, Valentin; Stufler, Benedikt (2020). A decorated tree approach to random permutations in substitution-closed classes. *Electronic Journal of Probability*, 25:online.

DOI: <https://doi.org/10.1214/20-ejp469>



Electron. J. Probab. **25** (2020), article no. 67, 1–52.  
ISSN: 1083-6489 <https://doi.org/10.1214/20-EJP469>

## A decorated tree approach to random permutations in substitution-closed classes

Jacopo Borga\*

Mathilde Bouvel<sup>†</sup>  
Benedikt Stufler<sup>§</sup>

Valentin Féray<sup>‡</sup>

### Abstract

We establish a novel bijective encoding that represents permutations as forests of decorated (or enriched) trees. This allows us to prove local convergence of uniform random permutations from substitution-closed classes satisfying a criticality constraint. It also enables us to reprove and strengthen permutation limits for these classes in a new way, that uses a semi-local version of Aldous' skeleton decomposition for size-constrained Galton–Watson trees.

**Keywords:** permutation patterns; local and scaling limits; Galton–Watson trees.

**AMS MSC 2010:** 60C05; 05A05.

Submitted to EJP on April 15, 2019, final version accepted on May 26, 2020.

Supersedes arXiv:1904.07135.

## 1 Introduction

### 1.1 Uniform random permutations in classes: some background and overview of our results

We assume some familiarity of the reader with basic definitions of permutation patterns and permutation classes, *i.e.*, what is a pattern, an occurrence and a consecutive occurrence, a class, its basis, . . . If needed, the definitions of these notions are given at the end of the introduction.

Permutation classes are classically studied from an enumerative point of view, *i.e.*, one wants to compute the number of permutations of any fixed size in a given class or the generating function of the class (possibly refining according to some statistics). In recent years, there has also been an increasing interest in the behaviour of a large typical permutation taken in a given permutation class. We refer for example to [13, 24, 25, 26, 32, 33, 43, 44] for results on random  $\tau$ -avoiding permutations with  $\tau$  of size 3. Other specific classes (or sets of permutations) have been studied: permutations avoiding a

\*Institut für Mathematik, Universität Zürich. E-mail: [jacopo.borga@math.uzh.ch](mailto:jacopo.borga@math.uzh.ch)

<sup>†</sup>Institut für Mathematik, Universität Zürich. E-mail: [mathilde.bouvel@math.uzh.ch](mailto:mathilde.bouvel@math.uzh.ch)

<sup>‡</sup>Institut für Mathematik, Universität Zürich. E-mail: [valentin.feray@math.uzh.ch](mailto:valentin.feray@math.uzh.ch)

<sup>§</sup>Institut für Mathematik, Universität München. E-mail: [stufler@math.lmu.de](mailto:stufler@math.lmu.de)

monotone pattern of any size [27], separable permutations [9], square permutations [14, 15], doubly alternating Baxter permutations [20]. The recent paper [7] by Bassino, Bouvel, Féray, Gerin, and Maazoun uses singularity analysis methods to study random permutations from *substitution-closed* classes satisfying some analytic assumptions.

The present work takes a probabilistic approach to the analysis of random permutations from substitution-closed classes. We establish a novel encoding of these permutations as decorated conditioned monotype Galton–Watson forests, hence integrating them into the framework of random enriched trees and tree-like structures introduced by Stufler [48, 50]. This yields a unified and powerful way for describing their asymptotic shape on a global and local scale.

- As a first application we give a new proof of the main scaling limit result of [7, 9] by using an extension of Aldous’ skeleton decomposition (see [6] for Aldous’ original statement, and Theorem 4.2 for our extension). This new proof works under weaker conditions and makes transparent the connection to random trees which was suggested, but unclear, in [7] (see in particular Remark 1.11 or the beginning of Section 1.7 there). In particular, our proof yields a probabilistic interpretation of the conditions under which this scaling limit result holds (see Section 1.6).
- Our second main contribution is a novel quenched local limit for random permutations from substitution-closed classes. Here we use fringe subtree count asymptotics and the skeleton decomposition to describe a concentration phenomenon for consecutive patterns. This notion of convergence has recently been introduced by Borga in [13], where such limits were proven for random permutations avoiding patterns of length 3.

The rest of the introduction defines substitution-closed classes and provides details on our results and on the approach used in this paper.

## 1.2 Substitution of permutations and closed classes

To define the substitution operation, it is convenient to think of permutations as diagrams. That is, if  $n$  denotes the size of a permutation  $\nu$ , we may identify  $\nu$  with the set of points  $(i, \nu(i))$  (for  $i$  in  $[n]$ ). The substitution  $\theta[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $\theta, \nu^{(1)}, \dots, \nu^{(d)}$  are permutations and  $d$  is the size of  $\theta$ , is then obtained as follows. For each  $i$ , we first replace the point  $(i, \theta(i))$  with the diagram of  $\nu^{(i)}$ . Then rescaling the rows and columns yields the diagram of a bigger permutation, which is by definition  $\theta[\nu^{(1)}, \dots, \nu^{(d)}]$ . A permutation of size greater than 2 is called *simple* if it cannot be obtained as the substitution of smaller permutations. An example of substitution is given in Fig. 1.

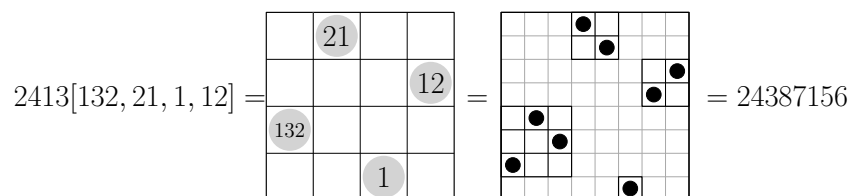


Figure 1: Example of substitution of permutations

As said above, this article considers *substitution-closed classes* of permutations, i.e., classes  $\mathcal{C}$  such that  $\theta, \nu^{(1)}, \dots, \nu^{(d)} \in \mathcal{C}$  implies  $\theta[\nu^{(1)}, \dots, \nu^{(d)}] \in \mathcal{C}$ . Alternatively, a class is substitution-closed if and only if its basis (i.e., the avoided patterns defining the class) consists only of simple permutations. In particular, there are uncountably many substitution-closed permutation classes. Due to their nice combinatorial structure (see

Section 2), substitution-closed permutation classes are a nice general framework, where to investigate the properties of uniform random elements.

We note that a substitution-closed class  $\mathcal{C}$  is entirely determined by the set  $S$  of simple permutations in it (see Theorem 2.11). We consider this set  $S$  as the data of our problem, and the goal is, under various conditions on  $S$  to obtain convergence results for uniform random permutations in the class  $\mathcal{C}$ . These conditions will typically be expressed in terms of the generating functions of  $S$ , that we conveniently also denote  $S$ . From Stanley-Wilf-Marcus-Tardős' theorem [41], it always has a positive radius of convergence  $\rho_S > 0$  (except in the trivial case where  $\mathcal{C}$  is the set  $\mathfrak{S}$  of all permutations, which we exclude from now on; permutation classes different from  $\mathfrak{S}$  are called *proper*).

### 1.3 Permuton convergence of substitution-closed classes

The notion of permutons was introduced in [29] to describe limits of permutation sequences. Formally, a permuton is a probability measure on the unit square  $[0, 1]^2$ , whose projection on each axis is the Lebesgue measure on  $[0, 1]$  (we say that the measure has uniform marginals). Permutations can be seen as permutons by considering the rescaled diagrams; we will denote  $\mu_\nu$  the permuton associated with the permutation  $\nu$ . The weak topology on measures gives then a natural meaning to the convergence of a sequence of permutations to a given permuton. A nice feature is that the convergence in terms of permutons is equivalent to the *convergence of pattern proportions*. We refer to [7, Section 2] for details.

Some specific permutons have been described as limits of permutation classes, as in [12, Chapter 6], [15] and [9, 7, 8]. Among these, the *biased Brownian separable permuton*  $\mu^{(p)}$  of parameter  $p$  is a random permuton, constructed from a Brownian excursion and independent signs associated with its local minima, see Maazoun [40]. It was proved in [7, 9] that this is a universal limiting object for substitution-closed permutations classes, in the sense that uniform random permutations in many substitution-closed classes converge to  $\mu^{(p)}$ , for some  $p$ . In this article, we give a new proof of this theorem that is based on an extension of Aldous' skeleton decomposition [6] and the framework of random enriched trees and tree-like structures [48, 50].

**Theorem 1.1.** *Let  $\nu_n$  be the uniform  $n$ -sized permutation from a proper substitution-closed class of permutations  $\mathcal{C}$ . Suppose that*

$$S'(\rho_S) > \frac{2}{(1 + \rho_S)^2} - 1, \quad (1.1)$$

or

$$S'(\rho_S) = \frac{2}{(1 + \rho_S)^2} - 1 \quad \text{and} \quad S''(\rho_S) < \infty. \quad (1.2)$$

Then

$$\mu_{\nu_n} \xrightarrow{d} \mu^{(p)},$$

with  $\mu^{(p)}$  denoting the biased Brownian separable permuton with an explicit parameter given by Eq. (5.1) page 32. This includes the case of uniform separable permutations, for which  $S = \emptyset$  and  $p = 1/2$ .

Specifically, the result [9, Thm. 1.2] corresponds to the special case where  $S = \emptyset$ , and [7, Thm. 1.10] corresponds to the special case where Eq. (1.1) is satisfied. The result [7, Thm. 7.8] corresponds to the case where Eq. (1.2) is satisfied and additionally  $S(z)$  is amenable to singularity analysis.

### 1.4 Local convergence: a concentration phenomenon for substitution-closed classes

In addition to scaling limits, our decorated tree approach also allows us to obtain local limit results for uniform random permutations in substitution-closed classes. For this, we use a local topology for permutations recently defined by Borga in [13]. This topology is the analogue of the celebrated Benjamini–Schramm convergence for graphs, in the sense that we look at the neighbourhood of a *random element* of the permutation. Pleasantly, convergence for this local topology is equivalent to the *convergence of consecutive pattern proportions*.

For convenience, we present our results in term of consecutive patterns. For a permutation  $\nu$  and a pattern  $\pi$ , we denote by  $c\text{-occ}(\pi, \nu)$  the number of consecutive occurrences of a pattern  $\pi$  in  $\nu$ ; for instance, for  $\pi = 21$  (resp.  $\pi = 321$ ), these are the number of descents (resp. double-descents) in the permutation.

**Theorem 1.2.** *Let  $\mathcal{C}$  be a proper substitution-closed permutation class and assume that*

$$\mathcal{S}'(\rho_S) \geq \frac{2}{(1 + \rho_S)^2} - 1. \quad (1.3)$$

*For each  $n \in \mathbb{N}$ , we consider a uniform random permutation  $\nu_n$  of size  $n$  in  $\mathcal{C}$ . Then, for each pattern  $\pi \in \mathcal{C}$ , there exists  $\gamma_{\pi, \mathcal{C}}$  in  $[0, 1]$  such that*

$$\frac{1}{n} c\text{-occ}(\pi, \nu_n) \xrightarrow{P} \gamma_{\pi, \mathcal{C}}. \quad (1.4)$$

We note that the hypothesis made in this theorem is slightly weaker than that for scaling limits. The theorem shows the convergence of all random variables  $\frac{1}{n} c\text{-occ}(\pi, \nu_n)$  to deterministic constants, revealing a “concentration” phenomenon in substitution-closed class under hypothesis (1.3). The constants  $\gamma_{\pi, \mathcal{C}}$  can be constructed from local limits of conditioned Galton–Watson trees around a random leaf, see Section 6 and in particular Theorem 6.22. They depend both on the pattern  $\pi$  and on the class  $\mathcal{C}$ .

### 1.5 Proof methodology

Start with a permutation  $\nu$  of size  $n \geq 2$ . If it is not simple nor monotone<sup>1</sup>, it can be written as  $\theta[\nu^{(1)}, \dots, \nu^{(d)}]$ , for some smaller permutations  $\theta, \nu^{(1)}, \dots, \nu^{(d)}$ . We can iterate this decomposition on  $\theta, \nu^{(1)}, \dots, \nu^{(d)}$ : as long as they are not simple nor monotone, we decompose them further through substitution. The result is a representation of  $\nu$  as a tree with  $n$  leaves, whose internal vertices are decorated by monotone or simple permutations. We call positive (resp. negative) a decoration with a monotone increasing (resp. decreasing) permutation. From a result of Albert and Atkinson [2], the decorated tree representation of a permutation is unique if we forbid the children of a vertex with a positive (resp. negative) decoration to have themselves a positive (resp. negative) decoration. The resulting tree is called the *canonical tree* of the permutation. Details on this construction, standard in the permutation pattern literature, are given in Section 2.2.

From a probabilistic point of view, one wants to consider the tree associated with a uniform random permutation in a class  $\mathcal{C}$  and possibly to recognize some standard tree models. One can show that, if the class is substitution-closed, the associated random canonical tree is a *multitype* random Galton–Watson tree with some specific offspring distribution conditioned on having  $n$  leaves. The need to have several types comes from the condition on positive and negative decorations: this forces us to consider children of

<sup>1</sup>By definition, for  $k \geq 2$ , there are exactly two monotone permutations of size  $k$ : the monotone increasing one  $12 \cdots k$  and the monotone decreasing one  $k \cdots 21$ .

positively and negatively decorated vertices to be of a different type from other vertices in the tree.

Results on conditioned multitype Galton–Watson trees do exist in the literature: in particular, there are some scaling limit results under finite or infinite variance assumptions [11, 18, 42], and local limit results around the root [1, 47] for such trees. Nevertheless these results do not cover our needs.

- For the scaling limit results on permutations, we need information on the type and outdegree of the closest common ancestors of randomly selected leaves (while tree scaling limit results only give information on the genealogy of such leaves).
- For the local limit results on permutations, we need some local limit results around a random leaf, and not around the root. For studying local convergence of random separable permutations we additionally require joint convergence with the parity of the height of the leaf.

We therefore do not use this encoding as multitype Galton–Watson trees, but rather provide a novel encoding of random permutations in substitution-closed classes as *decorated monotype Galton–Watson forests*. That is, random plane forests where each vertex is enriched with an independent local structure. This integrates the random permutations naturally into the framework of random tree-like structures [48].

To identify permutations with decorated forests, we first note that a generic permutation is the  $\oplus$ -sum of an ordered sequence of  $\oplus$ -indecomposable permutations, *i.e.*, of permutations which cannot be obtained as a substitution  $12[\pi^{(1)}, \pi^{(2)}]$  (see Theorem 2.5 below). We then associate to each of these  $\oplus$ -indecomposable permutations its canonical tree. To those trees, we apply a packing procedure. This packing procedure merges vertices decorated with a simple permutation with its children having a positive decoration. As a consequence, we do not need anymore to distinguish between positive and negative decorations. The resulting tree, called packed tree of the ( $\oplus$ -indecomposable) permutation, is still a decorated tree with  $n$  leaves, but the decorations are now more complicated objects than permutations, being themselves trees of permutations (called  $\mathcal{S}$ -gadget below, see Section 2.3 for details). The advantage of this new representation is that there is no condition on the decoration of a vertex, depending on the one of its parent. As a result of this construction, any permutation is represented as an ordered sequence of decorated trees, *i.e.*, an ordered decorated forest, without any constraint on the decorations (Theorem 2.19). We note that this representation is a bijection from the set of all permutations to ordered decorated forests, and could thus be of interest, independently from its application to the study of random elements in substitution-closed classes done here.

To study random permutations of size  $n$  taken uniformly at random in a substitution-closed class  $\mathcal{C}$ , we use a result on convergent Gibbs partitions (see Stufler [49, Thm. 3.1]) to prove that the associated ordered decorated forest contains a giant tree of size  $n - O_p(1)$  (Theorem 3.2 page 17). It is therefore enough to study a random decorated tree with  $n$  leaves. Such trees have the same distribution as a *monotype* Galton–Watson tree  $T_n^\xi$  with a specific offspring distribution  $\xi$  conditioned on having  $n$  leaves. We can therefore use results or techniques on monotype Galton–Watson trees, which are much more developed than in the multitype case.

- In particular, to find the scaling limit of our permutations, there are some results on the genealogy and the outdegree of common ancestors of randomly chosen vertices (this is implicit in the original paper of Aldous, see [6, Eq. (49)]). We will refer to this as *Aldous’ skeleton decomposition*. In this article we will need an extension of this, considering also local neighbourhood of the common ancestors

(being therefore *semi-local*) and allowing to condition on the number of leaves instead of the number of vertices. Theorem 4.2 provides a general result to this effect, allowing to condition on the number of vertices with arity in any given set  $\Omega \subseteq \mathbb{N}_0$  satisfying  $\mathbb{P}(\xi \in \Omega) > 0$ .

- The literature also contains results on the number of (extended) fringe subtrees of  $T_n^\xi$  (and related models) isomorphic to a given tree [3, 28, 31, 48, 51, 52]. When  $\xi$  is critical, such results may be translated to local limit results for  $T_n^\xi$ , pointed at a random leaf (see Theorem 6.13). We shall however need and will prove a slightly stronger result when  $\xi$  is critical and additionally has finite variance, taking also into account the parity of the height of the pointed leaf (see Theorem 6.20 page 46).

The last step of the proofs (both in the scaling and local limit cases) is to translate the results on the packed trees to results on the permutation  $\nu_n$  itself. A difficulty here arises from the identification of positive and negative decorations in the packing construction. To invert this construction, and recover the correct signs on the decorated trees, we need to determine the distance to the closest ancestor decorated with a simple permutation. When  $\mathcal{S} \neq \emptyset$ , this ancestor is at a stochastically bounded distance, so that this inversion procedure is still local. However, when  $\mathcal{S} = \emptyset$ , i.e., in the case of *separable permutations*, there is no such ancestor and we need to go all the way to the root to invert the packing construction. This creates an extra difficulty, that we overcome by using a local limit theorem for the length of “bones” in the skeleton decomposition.

## 1.6 Interpretation of the various assumptions on $\mathcal{S}$

Our assumptions on  $\mathcal{S}$  might seem artificial but they are in fact very natural, after having introduced the above representation of permutations as decorated conditioned Galton–Watson forests. Namely

- Eq. (1.3) is equivalent to the fact that the Galton–Watson tree model is critical;
- Eq. (1.1) asks in addition that the offspring distribution has small exponential moments;
- finally, Eq. (1.2) means that the offspring distribution has no exponential moments, but finite variance.

Such hypotheses are classical in the analysis of conditioned Galton–Watson trees, and give a probabilistic meaning to the conditions used in [7]. In terms of substitution-closed classes, the small exponential moment condition is satisfied for most classes in the literature, see the discussion in [7, Section 1.4]. Although general classes satisfying Eq. (1.3) but not Eq. (1.1) have also been studied in previous works [7, Sec. 7], we do not know at present whether such classes exist.

There is however at least one class not satisfying Eq. (1.3): the class  $\text{Av}(2413)$ . The packed forest associated with a uniform random permutation  $\nu_n$  in this class has the distribution of a decorated conditioned Galton–Watson forest with a *subcritical offspring distribution*. It will therefore contain with high probability a unique vertex with macroscopic degree (see [31, 34, 38, 52]). This vertex is decorated with a large simple permutation  $\alpha_n$  in the class and the scaling (resp. local) limit of  $\nu_n$  could be described if one knew that of  $\alpha_n$ . In the current state of the art, studying a uniform random *simple*  $\alpha_n$  in  $\text{Av}(2413)$  does not seem to be a simpler problem than the original one of studying  $\nu_n$ , hence this approach appears to be ineffective for  $\text{Av}(2413)$ .

## 1.7 Outline of the paper

The paper is organized as follows. The rest of the introduction sets up some notation. Section 2 presents the combinatorial construction used in this paper, that is the canonical tree and packed forest associated with a permutation. Section 3 identifies the packed forest associated with a uniform random permutation in a substitution-closed class as a conditioned monotype Galton–Watson forest. We also discuss the existence of a giant tree in such a forest. In Section 4, we state and prove our improvement of Aldous’ skeleton decomposition. The last two sections are devoted to the proofs of the main theorems: Section 5 for the scaling limit result (Theorem 1.1) and Section 6 for the local limit result (Theorem 1.2).

## 1.8 Permutation patterns and permutation classes: basic definitions and notation

We let  $\mathbb{N}_0 = \{0, 1, \dots\}$  denote the collection of non-negative integers and  $\mathbb{N} = \{1, 2, \dots\}$  the collection of strictly positive integers. For any  $n \in \mathbb{N}$ , we denote the set of permutations of  $[n] := \{1, 2, \dots, n\}$  by  $\mathfrak{S}^n$ . We write permutations of  $\mathfrak{S}^n$  in one-line notation as  $\nu = \nu(1)\nu(2)\dots\nu(n)$ . For a permutation  $\nu \in \mathfrak{S}^n$  the size  $n$  of  $\nu$  is denoted by  $|\nu|$ . We let  $\mathfrak{S} := \bigcup_{n \in \mathbb{N}} \mathfrak{S}^n$  be the set of finite permutations. We write sequences of permutations in  $\mathfrak{S}$  as  $(\nu_n)_{n \in \mathbb{N}}$ .

We will often view a permutation  $\nu$  as its diagram, which is (as said earlier – see also the right part of Fig. 1) the set of points of the Cartesian plane at coordinates  $(j, \nu(j))$ .

If  $x_1 \dots x_n$  is a sequence of distinct numbers, let  $\text{std}(x_1 \dots x_n)$  be the unique permutation  $\pi$  in  $\mathfrak{S}^n$  that is in the same relative order as  $x_1 \dots x_n$ , i.e.,  $\pi(i) < \pi(j)$  if and only if  $x_i < x_j$ . Given a permutation  $\nu \in \mathfrak{S}^n$  and a subset of indices  $I \subseteq [n]$ , let  $\text{pat}_I(\nu)$  be the permutation induced by  $(\nu(i))_{i \in I}$ , namely,  $\text{pat}_I(\nu) := \text{std}((\nu(i))_{i \in I})$ . For example, if  $\nu = 87532461$  and  $I = \{2, 4, 7\}$  then  $\text{pat}_{\{2,4,7\}}(87532461) = \text{std}(736) = 312$ .

Given two permutations,  $\nu \in \mathfrak{S}^n$  for some  $n \in \mathbb{N}$  and  $\pi \in \mathfrak{S}^k$  for some  $k \leq n$ , we say that  $\nu$  contains  $\pi$  as a *pattern* (and we write  $\pi \leq \nu$ ) if  $\nu$  has a subsequence of entries order-isomorphic to  $\pi$ , that is, if there exists a subset  $I \subseteq [n]$  such that  $\text{pat}_I(\nu) = \pi$ . Denoting  $i_1, i_2, \dots, i_k$  the elements of  $I$  in increasing order, the subsequence  $\nu(i_1)\nu(i_2)\dots\nu(i_k)$  is called an *occurrence* of  $\pi$  in  $\nu$ . In addition, we say that  $\nu$  contains  $\pi$  as a *consecutive pattern* if  $\nu$  has a subsequence of adjacent entries order-isomorphic to  $\pi$ , that is, if there exists an interval  $I \subseteq [n]$  such that  $\text{pat}_I(\nu) = \pi$ . Using the same notation as above,  $\nu(i_1)\nu(i_2)\dots\nu(i_k)$  is then called a *consecutive occurrence* of  $\pi$  in  $\nu$ . All along the article, for any integers  $a, b \in \mathbb{Z}$  (resp.  $n \in \mathbb{N}$ ), the interval  $[a, b]$  (any interval  $I \subseteq [n]$ ) is to be understood as an integer interval, i.e., an interval contained in  $\mathbb{Z}$ . For real numbers  $a \leq b$ , we use the same notation  $[a, b]$  to denote the interval  $[a, b] \subseteq \mathbb{R}$ .

**Example 1.3.** The permutation  $\nu = 1532467$  contains 1423 as a pattern but not as a consecutive pattern and 321 as consecutive pattern. Indeed  $\text{pat}_{\{1,2,3,5\}}(\nu) = 1423$  but no interval of indices of  $\nu$  induces the permutation 1423. Moreover,  $\text{pat}_{[2,4]}(\nu) = \text{pat}_{\{2,3,4\}}(\nu) = 321$ .

We say that  $\nu$  *avoids*  $\pi$  if  $\nu$  does not contain  $\pi$  as a pattern. We point out that the definition of  $\pi$ -avoiding permutations refers to patterns and not to consecutive patterns. Given a set of patterns  $B \subseteq \mathfrak{S}$ , we say that  $\nu$  *avoids*  $B$  if  $\nu$  avoids  $\pi$ , for all  $\pi \in B$ . We denote by  $\text{Av}^n(B)$  the set of  $B$ -avoiding permutations of size  $n$  and by  $\text{Av}(B) := \bigcup_{n \in \mathbb{N}} \text{Av}^n(B)$  the set of  $B$ -avoiding permutations of arbitrary size.

A *permutation class*  $\mathcal{C}$  is a set of permutations closed under the operation of pattern-containment (i.e., if  $\nu \in \mathcal{C}$  and  $\pi \leq \nu$  then  $\pi \in \mathcal{C}$ ). We recall that every permutation class can be rewritten as a family of pattern-avoiding permutations, i.e., for every permutation class  $\mathcal{C}$  there exists a set of patterns  $B \subseteq \mathfrak{S}$  such that  $\mathcal{C} = \text{Av}(B)$ . Note that if one



permutation of  $B$  is contained in another then we may remove the larger one without changing the family. Thus we may take  $B$  to be an *antichain*, meaning that no element of  $B$  contains any others. In the case that  $B$  is an antichain we call it the *basis* of this family. We note that the basis of a class may be finite or infinite.

### 1.9 Probabilistic notation

In order to avoid any confusion, we write random quantities using **bold characters** to distinguish them from deterministic quantities. Moreover, given a random variable  $\mathbf{X}$ , we denote with  $\mathcal{L}(\mathbf{X})$  its law. Unless otherwise stated, all limits are taken as  $n \rightarrow \infty$ . Given a sequence of random variables  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  we write  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$  to denote convergence in distribution and  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$  to denote convergence in probability. We let  $O_p(1)$  represent an unspecified random variable  $\mathbf{Y}_n$  of a stochastically bounded sequence  $(\mathbf{Y}_n)_n$ .

Besides, the indicator of an event  $A$  is denoted  $\mathbb{1}[A]$ . Finally, the expression *with high probability* means *with probability tending to 1* (without precision on the speed of convergence).

### 1.10 Index of notation

Table 1 summarizes some notational conventions and frequently used terminology in this paper. In general, for a combinatorial class denoted by a curly letter, e.g.  $\mathcal{P}$ , we use the same letter  $\mathcal{P}(z)$  for its generating series,  $\rho_{\mathcal{P}}$  for the radius of convergence of  $\mathcal{P}(z)$ , a standard uppercase letter  $P$  for an object in the class, and a lowercase letter  $p_n$  with index  $n \geq 0$  for the number of objects of size  $n$ . Bijections between classes will be denoted by two upper case letters, e.g. CT (page 11) building the canonical tree of a permutation.

Table 1: Table of the main notation.

$\mathfrak{S}^n$	the set of permutations of size $n$ , page 7
$\mathcal{C}$	a substitution-closed class of permutations, page 12
$\mathcal{S}$	the subset of simple permutations in $\mathcal{C}$ , page 12
$\mathcal{T}$	the class of canonical trees associated with $\mathcal{C}$ , page 12
$\mathcal{P}$	the class of packed trees associated with $\mathcal{C}$ , page 14
$\nu_n$	the uniform random $n$ -sized permutation from $\mathcal{C}$ , page 16
$P_n = (T_n, \lambda_{T_n})$	the uniform random packed tree with $n$ leaves, page 19
$\mathfrak{T}^\bullet$	the collection of (possibly infinite) pointed plane trees, page 39
$\mathfrak{T}^{\bullet, \text{luf}}$	the collection of (possibly infinite) locally and upwards finite pointed plane trees, page 39
$\mathfrak{T}_{\mathcal{D}}^{\bullet, \text{luf}}$	the collection of (possibly infinite) locally and upwards finite decorated pointed plane trees, page 40
$\mathfrak{P}^{\bullet, \text{luf}}$	the collection of (possibly infinite) locally and upwards finite pointed packed trees, page 41

## 2 A novel encoding of permutations as forests of decorated trees

In this section we show that any substitution-closed class of permutations may be bijectively encoded as a forest of trees decorated (or enriched) with local structures. This goal is achieved in Theorem 2.19, in Subsection 2.4.

## 2.1 Basics on combinatorial classes and decorated trees

In this paper, we only consider rooted (a.k.a. planted) plane trees; *plane* means that the children of a given vertex are endowed with a linear order. Throughout the paper, the *outdegree*  $d_T^+(v)$  (or  $d^+(v)$  when there is no ambiguity) of a vertex  $v$  in a tree  $T$  is the number of its children (which is sometimes called *arity* in other works). Note that it may be different from the graph-degree: the edge to the parent (if it exists) is not counted in the outdegree. We consider both finite and infinite trees. We say a tree is *locally finite*, if all its vertices have finite degree. A vertex of  $T$  is called a *leaf*, if it has outdegree zero. The collection of non-leaves (also called *internal vertices*) is denoted by  $V_{\text{int}}(T)$ . The *fringe subtree* of a tree  $T$  rooted at a vertex  $v$  is the subtree of  $T$  containing  $v$  and all its descendants. We will also speak of *branch* attached to a vertex  $v$  for a fringe subtree rooted at a child of  $v$ .

Any plane tree may be encoded in a canonical way as a subtree of the Ulam–Harris tree  $\mathcal{U}_\infty$ . The vertex set of  $\mathcal{U}_\infty$  is given by the collection of all finite sequences of positive integers, and the offspring of a vertex  $(i_1, \dots, i_k)$  is given by all sequences  $(i_1, \dots, i_k, j)$ ,  $j \geq 1$ . The root of  $\mathcal{U}_\infty$  is the unique sequence of length 0.

Moreover, most trees considered here carry some additional structures on their vertices from a *combinatorial class*. Let  $\mathcal{D}$  be a set and  $\text{size} : \mathcal{D} \rightarrow \mathbb{N}_0$  be a map from  $\mathcal{D}$  to the set of non-negative integers, associating to each object in  $\mathcal{D}$  its size. We say  $\mathcal{D}$  is an (unlabelled) combinatorial class, if for any  $n \in \mathbb{N}_0$  the number  $d_n$  of  $n$ -sized objects in  $\mathcal{D}$  is finite. This allows us to form the *generating series*

$$\mathcal{D}(z) = \sum_{n \geq 0} d_n z^n. \quad (2.1)$$

Note that we use the same curvy letter  $\mathcal{D}$  for the class and its generating series. This should hopefully not lead to confusions. Two combinatorial classes  $\mathcal{D}_1, \mathcal{D}_2$  are considered *isomorphic* if there is a size-preserving bijection between the two, or equivalently if they have the same generating series.

Various standard operations are available for combinatorial classes. For example, whenever  $\mathcal{D}$  has no objects of size 0, we can form the combinatorial class  $\text{Seq}(\mathcal{D})$ , which is the collection of finite sequences of objects from  $\mathcal{D}$ . The size of such a sequence is defined to be the sum of sizes of its components. We may also consider the subclass  $\text{Seq}_{\geq 1}(\mathcal{D}) \subset \text{Seq}(\mathcal{D})$  of non-empty sequences.

**Definition 2.1.** Let  $\mathcal{D}$  be a combinatorial class. A  $\mathcal{D}$ -decorated (or  $\mathcal{D}$ -enriched) tree is a rooted locally finite plane tree  $T$ , equipped with a function  $\text{dec} : V_{\text{int}}(T) \rightarrow \mathcal{D}$  from the set of internal vertices of  $T$  to  $\mathcal{D}$  such that the following holds: for each  $v$  in  $V_{\text{int}}(T)$ , the outdegree of  $v$  is exactly  $\text{size}(\text{dec}(v))$ .

This is a (planar) variant of Labelle’s *enriched trees* [39], which have been studied in [50, 48] from a probabilistic viewpoint.

## 2.2 Substitution decomposition and canonical trees

We recall classical concepts related to permutation classes, including for expository purposes the concepts sketched in the introduction.

**Definition 2.2.** Let  $\theta = \theta(1) \cdots \theta(d)$  be a permutation of size  $d$ , and let  $\nu^{(1)}, \dots, \nu^{(d)}$  be  $d$  other permutations. The substitution of  $\nu^{(1)}, \dots, \nu^{(d)}$  in  $\theta$ , denoted by  $\theta[\nu^{(1)}, \dots, \nu^{(d)}]$ , is the permutation of size  $|\nu^{(1)}| + \dots + |\nu^{(d)}|$  obtained by replacing each  $\theta(i)$  by a sequence of integers isomorphic to  $\nu^{(i)}$  while keeping the relative order induced by  $\theta$  between these subsequences.

Examples of substitution (see Fig. 2 below) are conveniently presented representing permutations by their diagrams: the diagram of  $\nu = \theta[\nu^{(1)}, \dots, \nu^{(d)}]$  is obtained by

inflating each point  $\theta(i)$  of  $\theta$  by a square containing the diagram of  $\nu^{(i)}$ . Note that each  $\nu^{(i)}$  then corresponds to a *block* of  $\nu$ , a block being defined as an interval of  $[\![\nu]\!]$  which is mapped to an interval by  $\nu$ .

Throughout this article, the increasing permutation  $12 \dots k$  will be denoted by  $\oplus_k$ , or even  $\oplus$  when its size  $k$  can be recovered from the context: this is the case in an inflation  $\oplus[\nu^{(1)}, \dots, \nu^{(d)}]$  where the size of  $\oplus$  is the number  $d$  of permutations inside the brackets. Similarly, we denote the decreasing permutation  $k \dots 21$  by  $\ominus_k$ , or  $\ominus$  when there is no ambiguity.

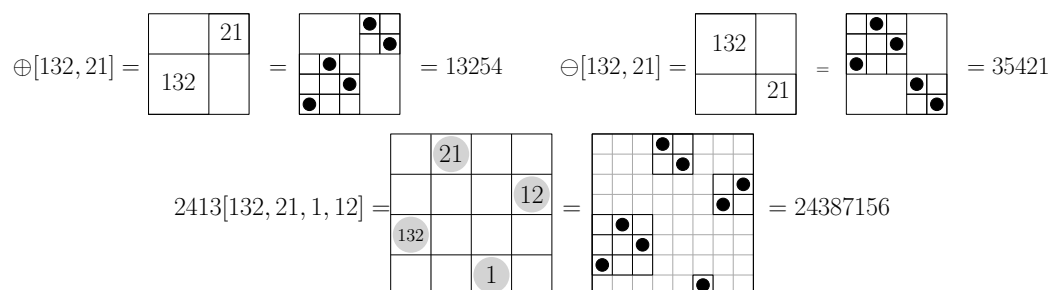


Figure 2: Substitution of permutations.

Permutations can be decomposed in a canonical way using recursively the substitution operation. To explain this, we first need to define several notions of indecomposable objects.

**Definition 2.3.** A permutation  $\nu$  is  $\oplus$ -indecomposable (resp.  $\ominus$ -indecomposable) if it cannot be written as  $\oplus[\nu^{(1)}, \nu^{(2)}]$  (resp.  $\ominus[\nu^{(1)}, \nu^{(2)}]$ ).

A permutation of size  $n > 2$  is simple if it contains no nontrivial block, i.e., if it does not map any nontrivial interval (i.e., a range in  $[n]$  containing at least two and at most  $n - 1$  elements) onto an interval.

For example, 451326 is not simple as it maps the interval  $[3, 5]$  onto the interval  $[1, 3]$ . The smallest simple permutations are 2413 and 3142 (there is no simple permutation of size 3). We denote by  $\mathcal{S}_{\text{all}}$  the set of simple permutations.

**Remark 2.4.** Usually in the literature, the definition of a simple permutation requires  $n \geq 2$  instead of  $n > 2$ , so that 12 and 21 are considered to be simple. However, for decomposition trees, 12 and 21 do not play the same role as the other simple permutations, that is why we do not consider them to be simple.

**Theorem 2.5** (Decomposition of permutations). Every permutation  $\nu$  of size  $n \geq 2$  can be uniquely decomposed as either:

- $\alpha[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $\alpha$  is a simple permutation (of size  $d \geq 4$ ),
- $\oplus[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $d \geq 2$  and  $\nu^{(1)}, \dots, \nu^{(d)}$  are  $\oplus$ -indecomposable,
- $\ominus[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $d \geq 2$  and  $\nu^{(1)}, \dots, \nu^{(d)}$  are  $\ominus$ -indecomposable.

**Remark 2.6.** The above theorem is essentially Proposition 2 in [2], presented with a slightly different point of view. The decomposition according to Theorem 2.5 is obtained from the one of [2, Proposition 2] by merging maximal sequences of nested substitutions in 12 (resp. 21) into a substitution in  $\oplus$  (resp.  $\ominus$ ). For example, the second item above for  $d = 4$  corresponds to  $12[\nu^{(1)}, 12[\nu^{(2)}, 12[\nu^{(3)}, \nu^{(4)}]]]$  with the notation of [2]. With this

obvious rewriting, the statements of [2, Proposition 2] and of Theorem 2.5 are trivially equivalent.

This decomposition theorem can be applied recursively inside the permutations  $\nu^{(i)}$  appearing in the items above, until we reach permutations of size 1. Doing so, a permutation  $\nu$  can be naturally encoded by a rooted labelled plane tree  $\text{CT}(\nu)$  as follows. (The notation  $\text{CT}(\nu)$  stands for *canonical tree* – see Theorem 2.7.)

- If  $\nu = 1$  is the unique permutation of size 1, then  $\text{CT}(\nu)$  is reduced to a single leaf.
- If  $\nu = \beta[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $\beta$  is either a simple permutation or the increasing (resp. decreasing) permutation (denoted by  $\oplus$ , resp.  $\ominus$ ), then  $\text{CT}(\nu)$  has a root of degree  $d$  labelled by  $\beta$  and the subtrees attached to the root are  $\text{CT}(\nu^{(1)}), \dots, \text{CT}(\nu^{(d)})$  (in this order from left to right).

From the above theorem, the decomposition  $\nu = \beta[\nu^{(1)}, \dots, \nu^{(d)}]$  exists and is unique if  $|\nu| \geq 2$ . Moreover,  $\nu^{(1)}, \dots, \nu^{(d)}$  have size smaller than  $\nu$  so that this recursive procedure always terminates and its result is unambiguously defined. The map  $\text{CT}$  is therefore well-defined. An example of this construction is shown on Fig. 3.

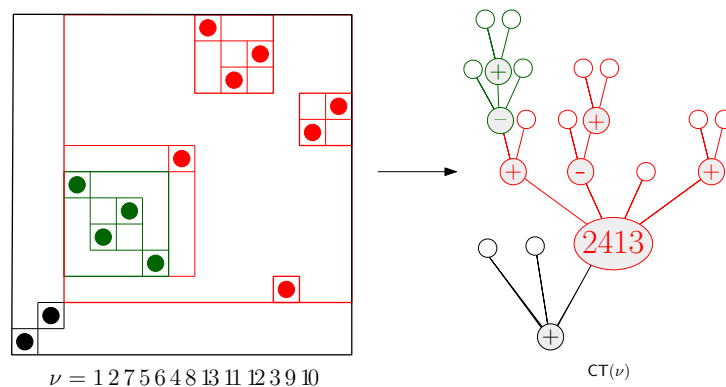


Figure 3: A permutation  $\nu$  and its decomposition tree  $\text{CT}(\nu)$ . To help the reader understand the construction, we have coloured accordingly some blocks of  $\nu$  and some subtrees of  $\text{CT}(\nu)$ .

Since the labels of the vertex record the permutation  $\beta$  in which we substitute, it is clear that  $\text{CT}$  is injective. Moreover, its inverse (once restricted to  $\text{CT}(\mathfrak{S})$ ) is immediate to describe, simply by performing the iterated substitutions recorded in the tree. We are just left with identifying the image set of  $\text{CT}$ . Recall that  $\mathcal{S}_{\text{all}}$  denotes the set of all simple permutations, and let  $\mathcal{M}$  be the set of all monotone (increasing or decreasing) permutations of size at least 2. Denote  $\widehat{\mathcal{S}}_{\text{all}} := \mathcal{S}_{\text{all}} \cup \mathcal{M}$ .

**Definition 2.7.** A canonical tree is an  $\widehat{\mathcal{S}}_{\text{all}}$ -decorated tree such that we cannot find two adjacent vertices both decorated with increasing permutations (i.e., with  $\oplus$ ) or both decorated with decreasing permutations (i.e., with  $\ominus$ ).

Canonical trees are also known in the literature under several names: decomposition trees, substitution trees, ... We choose the term *canonical* to be consistent with [7]. The following is an easy consequence of Theorem 2.5.

**Proposition 2.8.** The map  $\text{CT}$  defines a size-preserving bijection from the set of all permutations to the set of all canonical trees, the size of a tree being its number of leaves.

**Remark 2.9.** We note that the inverse map  $\text{CT}^{-1}$ , which builds a permutation from a canonical tree performing nested substitutions, can obviously be extended to all  $\widehat{\mathcal{S}}_{\text{all}}$ -decorated trees, regardless of whether they contain  $\oplus - \oplus$  or  $\ominus - \ominus$  edges. However,  $\text{CT}^{-1}$  is no longer injective on this larger class of “non-canonical” decomposition trees.

We will be interested in the restriction of  $\text{CT}$  to some permutation class. The following condition ensures that its image has a nice description.

**Definition 2.10.** A permutation class  $\mathcal{C}$  is substitution-closed if for every  $\theta, \nu^{(1)}, \dots, \nu^{(d)}$  in  $\mathcal{C}$  it holds that  $\theta[\nu^{(1)}, \dots, \nu^{(d)}] \in \mathcal{C}$ .

**Proposition 2.11.** Let  $\mathcal{C}$  be a substitution-closed permutation class, and assume<sup>2</sup> that  $12, 21 \in \mathcal{C}$ . Denote by  $\mathcal{S}$  the set of simple permutations in  $\mathcal{C}$ . The set of canonical trees encoding permutations of  $\mathcal{C}$  is the set of canonical trees with decorations in  $\widehat{\mathcal{S}} := \mathcal{S} \cup \mathcal{M}$ .

*Proof.* First, if a canonical tree contains a vertex decorated by a simple permutation  $\alpha \notin \mathcal{S}$ , then the corresponding permutation  $\nu$  contains the pattern  $\alpha \notin \mathcal{C}$ , and hence  $\nu \notin \mathcal{C}$ . Second, by induction, all canonical trees with decorations in  $\widehat{\mathcal{S}}$  encode permutations of  $\mathcal{C}$ , because  $\mathcal{C}$  is substitution-closed. If necessary, details can be found in [2, Lemma 11].  $\square$

### 2.3 Packed decomposition trees

From now until the end of the article we fix a substitution-closed class  $\mathcal{C}$  such that  $12, 21 \in \mathcal{C}$  and we denote with  $\mathcal{S}$  the set of simple permutations in  $\mathcal{C}$ . The assumption that we are working in  $\mathcal{C}$  rather than in the set of all permutations is however often tacit: for example, we simply refer to canonical trees instead of canonical trees with decorations in  $\widehat{\mathcal{S}} = \mathcal{S} \cup \mathcal{M}$ . We let  $\mathcal{T}$  denote the collection canonical trees with decorations in  $\widehat{\mathcal{S}}$ , and  $\mathcal{T}_{\text{not}\oplus} \subset \mathcal{T}$  the subset of canonical trees with a root that is **not** labelled  $\oplus$ .

In this section we introduce a new family of trees called “packed trees” and describe a bijection between the collection  $\mathcal{T}_{\text{not}\oplus} \subset \mathcal{T}$  and packed trees. Packed trees are decorated trees, whose decorations are themselves trees. Let us define these decorations, that we call *gadgets*.

**Definition 2.12.** An  $\mathcal{S}$ -gadget is an  $\widehat{\mathcal{S}}$ -decorated tree of height at most 2 such that:

- The root is an internal vertex decorated by a simple permutation;
- The children of the root are either leaves or decorated by an increasing permutation.

The size of a gadget is its number of leaves.

We denote with  $\mathcal{G}(\mathcal{S})$  the set of  $\mathcal{S}$ -gadgets. An example of size 7 is shown on Fig. 4.

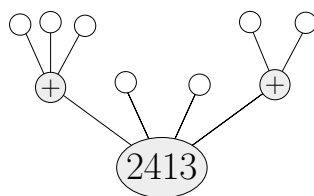


Figure 4: An  $\mathcal{S}$ -gadget.

Finally, let  $\widehat{\mathcal{G}(\mathcal{S})} = \mathcal{G}(\mathcal{S}) \cup \{\otimes_k, k \geq 2\}$ , where, for each integer  $k \geq 2$ , the object  $\otimes_k$  has size  $k$ . To shorten notation,  $\widehat{\mathcal{G}(\mathcal{S})}$  is sometimes denoted  $\mathcal{Q}$  in the following.

<sup>2</sup>Otherwise,  $\mathcal{C} \subseteq \{12 \dots k : k \geq 1\}$  or  $\mathcal{C} \subseteq \{k \dots 21 : k \geq 1\}$  and these cases are trivial.

**Definition 2.13.** An  $\mathcal{S}$ -packed tree is a  $\widehat{\mathcal{G}(\mathcal{S})}$ -decorated tree, its size being its number of leaves.

**Remark 2.14.** We will often refer to  $\mathcal{S}$ -packed trees simply as packed trees since in our analysis, the substitution-closed class  $\mathcal{C}$  and its set of simple permutations  $\mathcal{S}$  will be fixed.

An example of packed tree is shown on the right-hand side of Fig. 5.

**Remark 2.15.** Note that in Fig. 5 the subscript  $k$  is not reported in the vertices decorated by an element in  $\{\otimes_k, k \geq 2\}$ . Indeed it can be easily recovered by counting the number of children of the vertex.

We now describe a bijection between canonical trees with a root that is not labelled  $\oplus$  and packed trees. Given a tree  $T \in \mathcal{T}_{\text{not}\oplus}$  the corresponding packed tree  $\text{PA}(T)$  is obtained modifying  $T$  as follows.

- For each internal vertex  $v$  of  $T$  labelled by a simple permutation, we build an  $\mathcal{S}$ -gadget  $G_v$  whose internal vertices are  $v$  and the  $\oplus$ -children of  $v$ , the parent-child relation in  $G_v$  and the left-to-right order between children are inherited from the ones in  $T$ , and we add leaves so that the outdegree of each internal vertex is the same in  $G_v$  as in  $T$ . Then, in  $\text{PA}(T)$ , we merge  $v$  and the  $\oplus$ -children of  $v$  into a single vertex decorated by  $G_v$ .
- The remaining vertices of  $T$ , decorated by  $\ominus_k$  or  $\oplus_k$ , are decorated by  $\otimes_k$  instead.

An example is given on Fig. 5. As a preparation for the inversion procedure, let us note the following: if a vertex  $\tilde{v}$  in  $\text{PA}(T)$  has a decoration  $\otimes_k$  and his parent is decorated by an  $\mathcal{S}$ -gadget, then the corresponding vertex  $v$  in  $T$  had decoration  $\ominus_k$ . Indeed, a vertex decorated by  $\oplus$  which is the child of a vertex  $v$  labelled by a simple permutation is included in  $G_v$ , and canonical trees do not contain  $\oplus - \oplus$  edges.

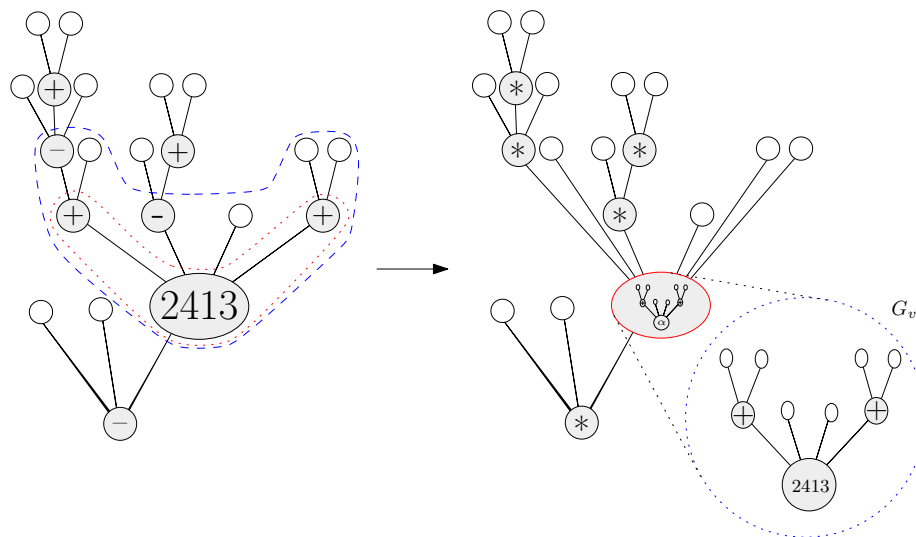


Figure 5: A canonical tree  $T$  with the corresponding packed tree  $\text{PA}(T)$ . The red dotted line highlights the vertices in  $T$  that are merged in the red vertex in  $\text{PA}(T)$ . The blue dashed line highlights the subtree in  $T$  that determines the decoration  $G_v$  in the tree  $\text{PA}(T)$ .

**Proposition 2.16.** The map  $\text{PA}$  defines a size-preserving bijection from the set  $\mathcal{T}_{\text{not}\oplus}$  of canonical trees with a root that is not labelled  $\oplus$  to the set of  $\mathcal{S}$ -packed trees.

*Proof.* We need just to show that the previous construction is invertible. Given a packed tree  $P$ , the corresponding tree  $T$  of  $\mathcal{T}_{\text{not}\oplus}$  (such that  $P = \text{PA}(T)$ ) is obtained by modifying  $P$  as follows.

- For each internal vertex  $\tilde{v}$  of  $P$  decorated by an  $\mathcal{S}$ -gadget  $G$ , we replace  $\tilde{v}$  by  $G$ , merging the leaves of  $G$  with the children of  $\tilde{v}$ , respecting their order. Namely, when doing this replacement, the root of the  $i$ -th subtree attached to  $\tilde{v}$  (from left to right) is merged with the  $i$ -th leaf of  $G$  (also from left to right).
- We replace each decoration  $\otimes_k$  with either  $\ominus_k$  or  $\oplus_k$ , with the following rule. If  $\tilde{v}$  is the root of  $P$  or the child of a vertex decorated by an  $\mathcal{S}$ -gadget, it receives label  $\ominus_k$ . Otherwise, if  $\tilde{v}$  is the child of a vertex also decorated by some  $\otimes$ , then we label  $\tilde{v}$  in the only way that prevents the creation of  $\oplus - \oplus$  or  $\ominus - \ominus$  edges.

This shows that  $\text{PA}$  defines a bijection.  $\square$

**Remark 2.17.** If  $T$  (or  $P$ ) is a tree with  $n$  leaves, we can label its leaves with number from 1 to  $n$  using a depth-first traversal of the tree from left to right. Then the  $i$ -th leaf of the canonical or packed tree associated to a permutation  $\nu$  corresponds to the  $i$ -th element in the one-line notation of  $\nu$ . We will use this identification between leaves and elements of the permutations later in the article.

## 2.4 Permutations are forests of decorated trees

Summing up the results obtained in the previous sections (in particular in Theorems 2.8, 2.11 and 2.16), we obtain a bijective encoding of  $\oplus$ -indecomposable permutations in  $\mathcal{C}$ :

**Lemma 2.18.** *The map*

$$\text{PA} \circ \text{CT} : \mathcal{C}_{\text{not}\oplus} \rightarrow \mathcal{P}$$

*is a size-preserving bijection from the set  $\mathcal{C}_{\text{not}\oplus}$  of all  $\oplus$ -indecomposable permutations in  $\mathcal{C}$  to the set  $\mathcal{P}$  of all  $\mathcal{S}$ -packed trees.*

By Theorem 2.5, any  $\oplus$ -decomposable permutation corresponds uniquely to a sequence of at least two  $\oplus$ -indecomposable permutations. Hence any permutation corresponds bijectively to a non-empty sequence of  $\oplus$ -indecomposable permutations. If we apply the bijection  $\text{PA} \circ \text{CT}$  to each we obtain a plane forest of packed trees. That is, it is an element of the collection  $\text{Seq}_{\geq 1}(\mathcal{P})$  of non-empty ordered sequences of  $\widehat{\mathcal{G}(\mathcal{S})}$ -decorated trees. We define the size of such a forest to be the total number of leaves. The function that maps a permutation of  $\mathcal{C}$  to the corresponding forest of packed trees is denoted by  $\text{DF}$  ( $\text{DF}$  stands for decorated forest). Summing up:

**Theorem 2.19.** *The function*

$$\text{DF} : \mathcal{C} \rightarrow \text{Seq}_{\geq 1}(\mathcal{P})$$

*is a size-preserving bijection between the substitution-closed class of permutations  $\mathcal{C}$  and the collection of forests of packed trees.*

## 2.5 Reading patterns in trees

Let us consider a permutation  $\nu$  in  $\mathcal{C}_{\text{not}\oplus}$  and the associated canonical and packed trees:  $T = \text{CT}(\nu)$  and  $P = \text{PA}(T)$ . Let  $I$  be a subset of  $[n]$ . Using Theorem 2.17,  $I$  can be seen as a subset of the leaves of  $T$  (or  $P$ ). The purpose of this section is to explain how to read out the pattern  $\pi = \text{pat}_I(\nu)$  on the trees  $T$  or  $P$ .

Let us first note that a pattern  $\pi = \pi(1) \dots \pi(k)$  is entirely determined when we know, for each  $i_1 < i_2$ , whether  $\pi(i_1)\pi(i_2)$  forms an *inversion* (i.e., an occurrence of the

pattern 21) or a non-inversion (occurrence of 12). Therefore, to read patterns on  $T$  (or  $P$ ), we should explain how to determine, for any two leaves  $\ell_1$  and  $\ell_2$  of  $I$ , whether the corresponding elements of  $\nu$  form an inversion or not (in the sequel, we will simply say that  $\ell_1$  and  $\ell_2$  form an inversion, and not refer anymore to the corresponding elements of  $\nu$ ).

Looking at  $T$ , this is rather easy. We consider the closest common ancestor of  $\ell_1$  and  $\ell_2$ , call it  $v$ . By definition,  $\ell_1$  and  $\ell_2$  are descendants of different children of  $v$ , say the  $i_1$ -th and  $i_2$ -th. Then the following holds:  $\ell_1$  and  $\ell_2$  form an inversion in  $\nu$  if and only if  $i_1$  and  $i_2$  form an inversion in the decoration  $\beta$  of  $v$ .

Let us now look at  $P$ . We consider the closest common ancestor  $u \in P$  of  $\ell_1$  and  $\ell_2$  and as before, we assume that  $\ell_1$  and  $\ell_2$  are descendants of the  $i_1$ -th and  $i_2$ -th children of  $u$ . Note that, in the packing bijection, the vertex  $u$  corresponds to  $v$  (the common ancestor of  $\ell_1$  and  $\ell_2$  in  $T$ ) potentially merged with other vertices.

Consider first the case that  $u$  is decorated by an  $\mathcal{S}$ -gadget  $G$ . Then  $G$  contains the information of the decoration of all vertices merged into  $u$ , including  $v$ . Therefore, whether  $\ell_1$  and  $\ell_2$  form an inversion in  $\nu$  can be determined by looking at the  $i_1$ -th and  $i_2$ -th leaves of the gadget  $G$  (see the example below).

If on the contrary  $u$  is not decorated by an  $\mathcal{S}$ -gadget but by a  $\otimes$ , we need to determine whether  $v$  is decorated with  $\oplus$  (implying that  $\ell_1$  and  $\ell_2$  form a non-inversion) or  $\ominus$  (resp., an inversion).

Assume first that there is a closest ancestor  $u'$  of  $u$  that is decorated with an  $\mathcal{S}$ -gadget. In this case, we claim that  $v$  is decorated by  $\ominus$  if  $d(u, u')$  is odd, and it is decorated by  $\oplus$  if  $d(u, u')$  is even. Indeed, decorations  $\oplus$  and  $\ominus$  alternate, and, by construction of the packing bijection, the vertex just above an  $\mathcal{S}$ -gadget is decorated by a  $\ominus$ .

It remains to analyse the case where  $u$  is decorated by  $\otimes$ , as well as all vertices on the path from  $u$  to the root  $r$  of  $P$ . By construction, this implies that the root of  $T$  is decorated by  $\ominus$ . So, using again the alternation of  $\oplus$  and  $\ominus$  in  $T$ , the decoration of  $v \in T$  is  $\ominus$  if  $d(u, r)$  is even, and  $\oplus$  if  $d(u, r)$  is odd.

We note in particular that the pattern induced by a set  $I$  of leaves in  $P$  is determined by any fringe subtree containing all leaves of  $I$  and rooted at any vertex decorated with an  $\mathcal{S}$ -gadget.

**Example 2.20.** Let  $\nu = 13\,12\,5\,3\,4\,2\,6\,11\,9\,10\,1\,7\,8$  be a permutation in  $\mathcal{C}_{\text{not}\oplus}$  with associated canonical and packed trees  $T = \text{CT}(\nu)$  and  $P = \text{PA}(T)$  shown in Fig. 6. We explain in the following example how to read out in  $P$  the pattern induced by the leaves  $\ell_1$ ,  $\ell_2$  and  $\ell_3$ .

The closest common ancestor  $u \in P$  of  $\ell_1$  and  $\ell_2$  is decorated with a  $\otimes$ , which is at distance 1 from its closest ancestor decorated with an  $\mathcal{S}$ -gadget. We can conclude that the leaves  $\ell_1$  and  $\ell_2$  induce an inversion (the closest ancestor  $v$  of  $\ell_1$  and  $\ell_2$  in  $T$  carries a  $\ominus$  decoration).

Now consider  $\ell_1$  and  $\ell_3$ . Their closest common ancestor  $u'$  in  $P$  is decorated with an  $\mathcal{S}$ -gadget. Note  $\ell_1$  and  $\ell_3$  are descendants of the first and fifth children of this  $\mathcal{S}$ -gadget; the corresponding leaves of the  $\mathcal{S}$ -gadget have the vertex decorated by 2413 as common ancestor and are attached to the branches corresponding to 2 and 3. We deduce that  $\ell_1$  and  $\ell_3$  do not form an inversion in  $\nu$ . Similarly,  $\ell_2$  and  $\ell_3$  do not form an inversion either in  $\nu$ .

Putting all together, the pattern induced by  $\ell_1$ ,  $\ell_2$  and  $\ell_3$  is 213. Let us check that it is indeed the case, by reading this pattern on the permutation. These three leaves correspond to the 4th, 6th and 12th elements of the permutation respectively, which have values 3, 2 and 7. The induced pattern is indeed 213.



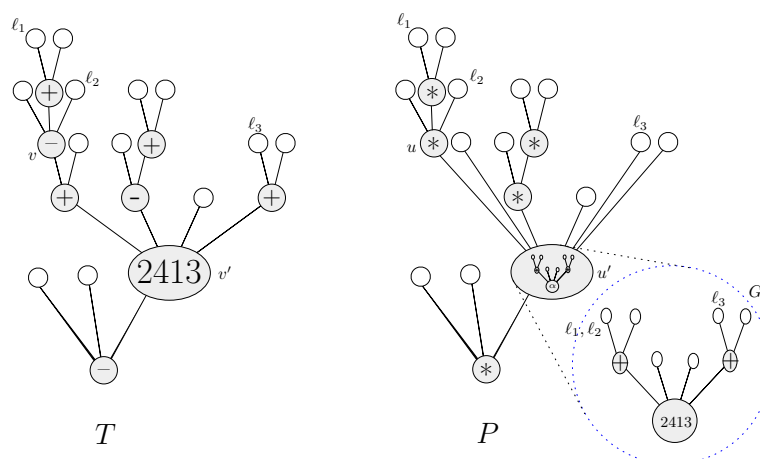


Figure 6: Reading patterns from trees – see Theorem 2.20

### 3 Random permutations and conditioned Galton–Watson trees

Throughout this section and the rest of the paper we assume that  $\mathcal{C}$  is a proper substitution-closed class of permutations, that is we exclude the case where  $\mathcal{C}$  is the class of all permutations. To avoid trivial cases, we furthermore assume that  $12, 21 \in \mathcal{C}$ .

Theorem 2.19 allows us to see a uniform random permutation  $\nu_n$  of size  $n$  in the substitution-closed permutation class  $\mathcal{C}$  as a uniform random forest of packed trees with  $n$  leaves. In the present section we apply Gibbs partition methods [49] to show that a giant component with size  $n - O_p(1)$  emerges, and the small fragments admit a limit distribution. This goal is achieved in Proposition 3.2. Since the size of the small fragments is stochastically bounded, this reduces the study of  $\nu_n$  to that of a uniform random packed tree with  $n$  vertices. The strength of this approach is that we do not need to make any additional assumptions on the class  $\mathcal{C}$ .

#### 3.1 Enumerative observations

Theorem 2.19 implies that the generating series of the class  $\mathcal{C}$  satisfies

$$\mathcal{C}(z) = \frac{\mathcal{P}(z)}{1 - \mathcal{P}(z)}. \quad (3.1)$$

From the definition of packed trees, we deduce the following equation for their generating series:

$$\mathcal{P}(z) = z + \mathcal{Q}(\mathcal{P}(z)), \quad (3.2)$$

where  $\mathcal{Q}(u) = \widehat{\mathcal{G}(\mathcal{S})}(u)$  is defined as the generating function of  $\widehat{\mathcal{G}(\mathcal{S})}$ . Via basic algebraic manipulations, we rewrite this as

$$\mathcal{P}(z) = z \left( 1 + \frac{\mathcal{Q}(\mathcal{P}(z))}{\mathcal{P}(z) - \mathcal{Q}(\mathcal{P}(z))} \right) = z \left( \frac{\mathcal{P}(z)}{\mathcal{P}(z) - \mathcal{Q}(\mathcal{P}(z))} \right) = z \mathcal{R}(\mathcal{P}(z)), \quad (3.3)$$

with

$$\mathcal{R}(u) = \frac{1}{1 - \mathcal{Q}(u)/u}. \quad (3.4)$$

By definition, an  $\mathcal{S}$ -gadget is described by a simple permutation of size say  $k$ , and  $k$  elements, which are either atoms (elements of size one) or increasing permutations of size at least two. Therefore

$$\mathcal{G}(\mathcal{S})(z) = \mathcal{S}\left(z + \frac{z^2}{1-z}\right) = \mathcal{S}\left(\frac{z}{1-z}\right),$$

and consequently,

$$\mathcal{Q}(z) = \widehat{\mathcal{G}(\mathcal{S})}(z) = \frac{z^2}{1-z} + \mathcal{S}\left(\frac{z}{1-z}\right). \quad (3.5)$$

Since we assumed that  $\mathcal{C}$  is proper, a celebrated result by Marcus and Tardos [41] states that the generating series  $\mathcal{C}(z)$  has positive radius of convergence. Hence the same holds for  $\mathcal{S}(z)$ , and consequently, for  $\mathcal{Q}(z)$  and  $\mathcal{R}(z)$ . A general result on solutions of implicit equations (such as (3.3)) [49, Lem. 3.3] implies that the  $n$ -th coefficient  $p_n$  of  $\mathcal{P}(z)$  satisfies the subexponentiality condition

$$\frac{p_n}{p_{n+1}} \rightarrow \rho_{\mathcal{P}} \quad \text{and} \quad \frac{1}{p_n} \sum_{i+j=n} p_i p_j \rightarrow 2\mathcal{P}(\rho_{\mathcal{P}}) < \infty, \quad (3.6)$$

as  $n \rightarrow \infty$ , with  $0 < \rho_{\mathcal{P}} < \infty$  denoting the radius of convergence of  $\mathcal{P}(z)$ . This even implies

$$\mathcal{P}(\rho_{\mathcal{P}}) < 1. \quad (3.7)$$

Indeed, if  $1 \leq \mathcal{P}(\rho_{\mathcal{P}}) < \infty$ , then there would exist a number  $0 \leq t \leq \rho_{\mathcal{P}}$  with  $\mathcal{P}(t) = 1$  and hence  $\widehat{\mathcal{G}(\mathcal{S})}(\mathcal{P}(t)) = \infty$  by (3.5). But this is not possible by Eqs. (3.3) and (3.4).

Eqs. (3.6) and (3.7) allow us apply [22, Thm. 4.8, 4.30] (or [17, Thm. 1]), yielding that the number  $c_n$  of  $n$ -sized permutations in  $\mathcal{C}$  satisfies

$$c_n \sim \frac{p_n}{(1 - \mathcal{P}(\rho_{\mathcal{P}}))^2}. \quad (3.8)$$

**Remark 3.1.** Eq. (3.2) identifies the class  $\mathcal{P}$  as so-called  $\mathcal{Q}$ -enriched parenthesizations. A classical bijection due to Ehrenborg and Méndez [21] consequently allows us to identify the class  $\mathcal{P}$  with the class of  $\mathcal{R}$ -enriched trees. The recursive equation  $\mathcal{P}(z) = z\mathcal{R}(\mathcal{P}(z))$  with  $\mathcal{R}$  given in Eq. (3.4) is actually a consequence of this general bijection.

### 3.2 A giant $\oplus$ -indecomposable component

Let  $\nu$  be a permutation in the proper substitution-closed class of permutations  $\mathcal{C}$ . From Theorem 2.5, we know that

- either  $\nu$  is  $\oplus$ -indecomposable,
- or  $\nu$  can be uniquely written as  $\nu = \oplus[\nu^{(1)}, \dots, \nu^{(d)}]$ , where  $d \geq 2$  and  $\nu^{(1)}, \dots, \nu^{(d)} \in \mathcal{C}_{\text{not}\oplus}$ , the set of  $\oplus$ -indecomposable permutations of  $\mathcal{C}$ .

In the first case, we set  $d = 1$  and  $\nu^{(1)} = \nu$  for convenience. Recall that Theorem 2.18 allows us to identify the classes  $\mathcal{C}_{\text{not}\oplus}$  and  $\mathcal{P}$ . The subexponentiality condition (3.6) allows us to apply the Gibbs partition result [49, Thm. 3.1] to obtain the following result (only the first part will be useful in this paper, but we state the whole version for completeness):

**Proposition 3.2.** *Let  $\nu_n$  be a uniform random permutation of size  $n$  in  $\mathcal{C}$  and define  $d, \nu^{(1)}, \dots, \nu^{(d)}$  as above. Let  $m$  be the smallest index such that  $|\nu^{(m)}| = \max(|\nu^{(1)}|, \dots, |\nu^{(d)}|)$ . Then  $\nu^{(m)}$  has size  $n - O_p(1)$ , and conditionally on its size,  $\nu^{(m)}$  is uniformly distributed among all  $|\nu^{(m)}|$ -sized  $\oplus$ -indecomposable permutations in  $\mathcal{C}$ .*

Moreover, the other components converge jointly in distribution:

$$((\nu^{(1)}, \dots, \nu^{(m-1)}), (\nu^{(m+1)}, \dots, \nu^{(d)})) \xrightarrow{d} ((\tilde{\nu}_1, \dots, \tilde{\nu}_{G_1}), (\tilde{\nu}_1, \dots, \tilde{\nu}_{G_2}))$$

with  $G_1, G_2$  denoting i.i.d. geometric random variables with distribution

$$\mathbb{P}(G_i = k) = \mathcal{P}(\rho_{\mathcal{P}})^k (1 - \mathcal{P}(\rho_{\mathcal{P}})), \quad k \geq 0,$$

and  $\bar{\nu}_i, \tilde{\nu}_i, i \geq 1$ , denoting independent copies of a Boltzmann-distributed random object  $\nu$  with distribution given by

$$\mathbb{P}(\nu = \nu) = \rho_{\mathcal{P}}^{|\nu|} / \mathcal{P}(\rho_{\mathcal{P}}).$$

**Remark 3.3.** We excluded the case of uniform unrestricted  $n$ -sized permutations. In this case, it is well-known that the permutation is with high probability  $\oplus$ -indecomposable. This follows for example from [48, Cor. 6.19] in the tree literature or from [16, Thm 3.4] in the permutation literature.

### 3.3 From permutations to simply generated trees

Proposition 3.2 and Lemma 2.18 reduce the study of the proper substitution-closed class  $\mathcal{C}$  to the study of the class  $\mathcal{P}$  of packed trees. In this section, we explain how a random tree in  $\mathcal{P}$  can be seen as a random simply generated tree with random decorations. This result may be seen as a special case of a sampling procedure [48, Sec. 6.4] for general enriched trees with a fixed number of leaves (so called enriched Schröder parenthesizations), but we present it in our specific setting to make the article more self-contained.

We can describe a packed tree  $P$  as a pair  $(T, \lambda_T)$  where  $T$  is a rooted plane tree and  $\lambda_T$  is a map from the internal vertices of  $T$  to the set  $\mathcal{Q} = \widehat{\mathcal{G}(\mathcal{S})}$  which records the decorations of the vertices.

In order to sample a uniform packed tree with  $n$  leaves, we first simulate a random rooted plane tree  $T_n$  and then a random decoration map  $\lambda_{T_n}$  as follows.

Define the weight-sequence  $\vec{q} = (q_k)_{k \geq 0}$ , where, for  $k \geq 2$ ,  $q_k$  denotes the  $k$ -th coefficient of the generating series  $\mathcal{Q}(z) = \widehat{\mathcal{G}(\mathcal{S})}(z)$ , while we set  $q_0 = 1$  and  $q_1 = 0$ . We consider the simply generated tree  $T_n$  (with  $n$  leaves) associated with weight-sequence  $\vec{q}$ , i.e., by definition,  $T_n$  is a random rooted plane tree such that

$$\mathbb{P}(T_n = T) = \frac{\prod_{v \in T} q_{d^+(v)}}{Z_n} = \frac{\prod_{v \in V_{\text{int}}(T)} q_{d^+(v)}}{Z_n}, \quad (3.9)$$

for all rooted plane trees  $T$  with  $n$  leaves (we recall that  $V_{\text{int}}(T)$  denotes the set of internal vertices of  $T$ ). Here,  $Z_n$  is the *partition function* given by

$$Z_n = \sum_T \prod_{v \in T} q_{d^+(v)},$$

where the sum runs over all rooted plane trees with  $n$  leaves. For a general introduction about simply generated trees see [31, Section 2.3].

Then, given a rooted plane tree  $T$ , let  $\lambda_T$  be the random map such that for all internal vertices  $v$  of  $T$ ,

$$\mathbb{P}(\lambda_T(v) = Q) = \frac{1}{q_{d^+(v)}} \quad \text{for all } Q \in \mathcal{Q} \text{ with } |Q| = d^+(v), \quad (3.10)$$

independently of all other choices. Namely, the decoration of each internal vertex  $v$  of  $T$  gets drawn uniformly at random among all  $d^+(v)$ -sized decorations in  $\mathcal{Q}$ , independently of all the other decorations.

**Lemma 3.4.** *The random packed tree  $P_n = (T_n, \lambda_{T_n})$  is uniform among all the packed trees with  $n$  leaves.*

*Proof.* Let  $P = (T, \lambda)$  be a packed tree with  $n$  vertices. Then

$$\begin{aligned} \mathbb{P}((T_n, \lambda_{T_n}) = (T, \lambda)) &= \mathbb{P}((T_n, \lambda_{T_n}) = (T, \lambda) | T_n = T) \cdot \mathbb{P}(T_n = T) \\ &= \left( \prod_{v \in V_{\text{int}}(T)} \frac{1}{q_{d_T^+(v)}} \right) \cdot \left( \frac{\prod_{v \in T} q_{d_T^+(v)}}{Z_n} \right) = \frac{1}{Z_n}, \end{aligned} \quad (3.11)$$

where in the second equality we use Eqs. (3.9) and (3.10).  $\square$

### 3.4 Random packed trees as conditioned Galton–Watson trees

Building on Theorem 3.4, in what follows we explain how to sample a uniform packed tree with  $n$  leaves as a randomly decorated Galton–Watson tree conditioned on having  $n$  leaves. Again, we refer to [48, Sec. 6.4] for a discussion in a more general context.

Let  $\rho_q$  denote the radius of convergence of the generating series  $\mathcal{Q}(z)$ . As observed in Section 3.1, it holds that  $\rho_q > 0$ . As we shall see, this implies that  $T_n$  has the distribution of a Galton–Watson tree conditioned of having  $n$  leaves, whose offspring distribution  $\xi$  is defined below (for similar discussion with fixed number of vertices, see [31, Section 4]).

The offspring distribution  $\xi$  is given by

$$\begin{cases} \mathbb{P}(\xi = 0) = a \\ \mathbb{P}(\xi = 1) = 0 \\ \mathbb{P}(\xi = k) = q_k t_0^{k-1} \text{ for } k \geq 2. \end{cases} \quad (3.12)$$

with  $a, t_0 > 0$  constants that are defined as follows. If  $\lim_{z \nearrow \rho_q} \mathcal{Q}'(z) \geq 1$ , let  $0 < t_0 \leq \rho_q$  be the unique number with  $\mathcal{Q}'(t_0) = 1$ . If the limit is less than 1, then set  $t_0 = \rho_q$ . Finally set  $a = 1 - \sum_{k \geq 2} q_k t_0^{k-1} > 0$ .

Note that the tilting in Eq. (3.12) previously appeared in [45, Proposition 2] (see also the discussion above Corollary 1 in the same paper).

We note that  $\xi$  is always aperiodic since  $q_k > 0$  for  $k \geq 2$  (because of the  $\circledast$  decorations). Moreover, we have

$$\mathbb{E}[\xi] = \mathcal{Q}'(t_0) \leq 1, \quad (3.13)$$

so that the Galton–Watson tree  $T^\xi$  of offspring distribution  $\xi$  is either subcritical or critical. It is a simple exercise to check that  $T^\xi$ , conditioned on having  $n$  leaves, has the same distribution as the simply generated tree  $T_n$  defined by Eq. (3.9).

To end this section, we characterize when this Galton–Watson tree model is critical. Below, we write  $\mathcal{S}'(\rho_S)$  for  $\lim_{z \nearrow \rho_S} \mathcal{S}'(z)$ , noting that this limit may be infinite.

**Proposition 3.5.** *It holds that  $\mathbb{E}[\xi] = 1$  if and only if*

$$\mathcal{S}'(\rho_S) \geq \frac{2}{(1 + \rho_S)^2} - 1. \quad (3.14)$$

*In this case,  $t_0 = \kappa/(1 + \kappa)$  for the unique number  $0 < \kappa \leq \rho_S$  with  $\mathcal{S}'(\kappa) = 2/(1 + \kappa)^2 - 1$ , and*

$$\mathbb{V}[\xi] = \kappa(1 + \kappa)^3 \mathcal{S}''(\kappa) + 4\kappa. \quad (3.15)$$

For the convenience of the reader, we note that the relation between  $t_0$  and  $\kappa$  can be rewritten as  $\kappa = \frac{t_0}{1 - t_0}$ .

*Proof.* It holds that

$$\mathcal{Q}'(z) = \frac{\mathcal{S}'(z/(1-z)) + z(2-z)}{(1-z)^2}$$

We perform the formally substitution  $z = y/(1+y)$  (which implies  $z = \rho_q \Leftrightarrow y = \rho_S$ ). This yields

$$\mathcal{Q}'(z) = (1+y)^2 \mathcal{S}'(y) + y^2 + 2y = 1 + (1+y)^2 \left[ \mathcal{S}'(y) + 1 - \frac{2}{(1+y)^2} \right]. \quad (3.16)$$

Recall that  $\mathbb{E}[\xi] = 1$  if and only if  $\lim_{z \nearrow \rho_q} \mathcal{Q}'(z) \geq 1$ . Since  $\rho_q = \frac{\rho_S}{1+\rho_S}$ , this shows the first part of the statement. The formula for  $t_0$  also follows from (3.16) and the definition of  $t_0$ . Finally, if  $\mathbb{E}[\xi] = 1$ , then

$$\begin{aligned} \mathbb{V}[\xi] &= \mathbb{E}[\xi(\xi-1)] = t_0 \mathcal{Q}''(t_0) \\ &= \kappa(1+\kappa)^3 \mathcal{S}''(\kappa) + 2\kappa(1+\kappa)^2 (\mathcal{S}'(\kappa) + 1) \\ &= \kappa(1+\kappa)^3 \mathcal{S}''(\kappa) + 4\kappa. \end{aligned} \quad \square$$

## 4 Semi-local convergence of the skeleton decomposition

The previous section establishes a connection between uniform random permutations and conditioned Galton–Watson trees. In this section, we provide a convergence result for skeletons induced by marked vertices in such trees. The application to permutations will be discussed in further sections.

Aldous [6, Eq. (49)] showed that the subtree spanned by a fixed number of random marked vertices in a large critical Galton–Watson tree admits a limit distribution. Here, we extend this *skeleton decomposition* so that it additionally describes the asymptotic local structure in  $o(\sqrt{n})$ -neighbourhoods around the marked vertices and their pairwise closest common ancestors. Note also that Aldous works with Galton–Watson trees conditioned on having  $n$  vertices, while we more generally consider Galton–Watson trees conditioned on having  $n$  vertices with out-degree in a given set  $\Omega$  (see [46] or [37] for scaling limit results under such conditioning).

### 4.1 Extracting the skeleton with a local structure

Let  $k \geq 1$  denote a fixed integer and  $T$  a (rooted) plane tree. We choose an ordered sequence  $\vec{v} = (v_1, \dots, v_k)$  of vertices in  $T$  (possibly with repetitions) that we call *marked vertices*. The goal of this section is to associate to this data some object recording:

- the genealogy between the marked vertices;
- the local structure around the *essential vertices* of  $T$ , which we define as the root of  $T$ , the marked vertices  $v_1, \dots, v_k$  and their pairwise closest common ancestors;
- the distances in the original tree between these vertices.

The reader can look at Fig. 7 to see the different steps of the construction.

The first step is to consider the subtree  $R(T, \vec{v})$  consisting of the vertices  $\vec{v}$  and all their ancestors. For each  $1 \leq j \leq k$  the vertex  $v_j$  in  $R(T, \vec{v})$  receives the label  $j$ . Note that the tree  $T$  may be constructed from the skeleton  $R(T, \vec{v})$  by attaching an ordered sequence of branches (rooted plane trees) at each corner of  $R(T, \vec{v})$ . Here we have to consider the corner below the root-vertex twice, since branches at this corner may either be added to the left or to the right of  $R(T, \vec{v})$ .

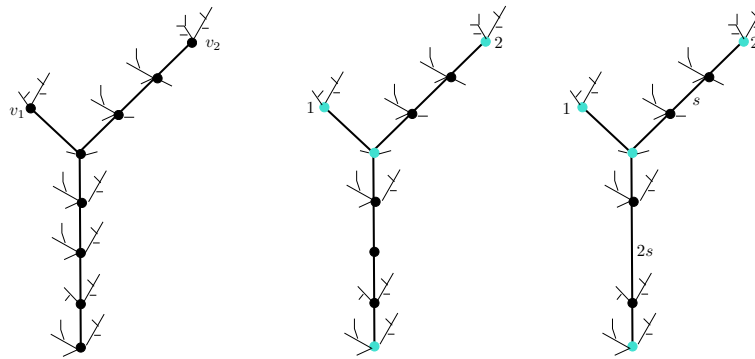


Figure 7: A tree  $T$  with two marked vertices  $v_1$  and  $v_2$ . In the left-most picture, the subtree  $R(T, \vec{v})$  is represented in bold, while branches attached to its corner are drawn with thinner lines. The middle picture represent  $R^{[1]}(T, \vec{v})$ : the essential vertices are in blue, and only one vertex of  $R(T, \vec{v})$  is at distance more than 1 from the closest essential vertex. The branches attached to that vertex do not belong to  $R^{[1]}(T, \vec{v})$ . The right-most picture represent  $s.R^{[1]}(T, \vec{v})$ . In particular, observe that the two middle edges of the path between the root and the branching vertex have been contracted into a single edge with label  $2s$ .

The second step is to remove the vertices of  $T$  which lie outside of the skeleton  $R(T, \vec{v})$  and are “far” from the essential vertices. For convenience, we call distance of any branch  $B$  (grafted on  $R(T, \vec{v})$ ) from a vertex  $w \in R(T, \vec{v})$  the distance in  $R(T, \vec{v})$  from  $w$  to the corner where  $B$  is attached. For any integer  $t \geq 0$ , we let  $R^{[t]}(T, \vec{v})$  denote the subtree of  $T$  that contains  $R(T, \vec{v})$  and all branches grafted on  $R(T, \vec{v})$  that have distance at most  $t$  from at least one essential vertex. In particular,  $R^{[t]}(T, \vec{v})$  contains all vertices of  $T$  that lie at distance at most  $t$  from the essential vertices.

The final step of the construction is to shrink the paths of  $R^{[t]}(T, \vec{v})$  consisting of the vertices whose attached branches have been removed in step 2. Indeed, we are interested in a scenario where the distance between any two essential vertices is much larger than  $2t$ . Consider two essential points  $x \neq y$  that are connected by a path not containing other essential vertices. Assume that  $x$  lies on the path from the root to  $y$ . If the distance between  $x$  and  $y$  is larger than  $2t$ , then the path joining  $x$  and  $y$  consists of a starting segment of length  $t$  that starts at  $x$ , a *middle segment* of positive length, and an end segment of length  $t$  that ends at  $y$ . By construction, the branches attached to inner vertices of the middle segment of  $R(T, \vec{v})$  do not appear in  $R^{[t]}(T, \vec{v})$ . For any real number  $s > 0$ , we let  $s.R^{[t]}(T, \vec{v})$  denote the result of contracting each middle segment to a single edge that receives a label given by the product of  $s$  and the number of deleted vertices in this segment.

## 4.2 The space of skeletons with a local structure

In the following, we will need to be more precise about the space in which  $s.R^{[t]}(T, \vec{v})$  lives and the topology we consider on it. In the above construction,  $s.R^{[t]}(T, \vec{v})$  is a tree with  $k$  distinguished vertices with outdegree in  $\Omega$ , where at most  $2k - 1$  edges have a (length-)label. Moreover, the distances between successive essential vertices are at most  $2t + 1$  (we say that two essential vertices are successive if the path going from one to the other does not contain any other essential vertex). The set of trees (without edge-labels) with  $k$  marked distinguished vertices with outdegree in  $\Omega$  such that the above distance condition holds is denoted  $\mathcal{T}_{k, \Omega}^{[t]}$ . Moreover, we say that  $G$  in  $\mathcal{T}_{k, \Omega}^{[t]}$  is *generic* if:

- there are  $2k$  distinct essential vertices (the root, the  $k$  distinguished vertices and  $k - 1$  closest common ancestors of pairs of distinguished vertices);
- the distances between successive essential vertices are exactly  $2t + 1$ .

We note that the edges with (length-)label are middle edges of the paths of length  $2t + 1$  between essential vertices, and hence depend only on the shape of the tree. We can therefore encode these labels as a vector in  $\mathbb{R}^{2k-1}$ , that has entries equal to 0 whenever the corresponding essential vertices are at distance  $2t$  or less. Finally,  $s.R^{[t]}(T, \vec{v})$  can be seen as an element of

$$\mathcal{T}_{k,\Omega}^{[t]} \times (\mathbb{R}_+)^{2k-1}.$$

(A similar identification is done by Aldous throughout the article [6].)

Using the discrete topology on  $\mathcal{T}_{k,\Omega}^{[t]}$  and the usual one on  $\mathbb{R}^{2k-1}$ , this gives a topology on  $\mathcal{T}_{k,\Omega}^{[t]} \times (\mathbb{R}_+)^{2k-1}$ , and then it makes sense to speak of convergence in distribution in this space. We can also speak of *density*, taking as reference measure the product of the counting measure on  $\mathcal{T}_{k,\Omega}^{[t]}$  and the Lebesgue measure on  $(\mathbb{R}_+)^{2k-1}$ . Finally we denote by Sh and Lab the natural projections from  $\mathcal{T}_{k,\Omega}^{[t]} \times (\mathbb{R}_+)^{2k-1}$  to  $\mathcal{T}_{k,\Omega}^{[t]}$  and  $(\mathbb{R}_+)^{2k-1}$ , respectively. In words Sh erases the labels and output the *shape* of the tree, while Lab outputs the vector of labels.

### 4.3 The limit tree

Throughout Section 4 we let  $T$  denote a (non-degenerate) critical Galton–Watson tree having an aperiodic offspring distribution  $\xi$ . We also assume that  $\xi$  has finite variance  $\sigma^2$ . We fix a subset  $\Omega \subseteq \mathbb{N}_0$  satisfying

$$\mathbb{P}(\xi \in \Omega) > 0. \quad (4.1)$$

Given a rooted plane tree  $T$ , we let  $|T|_\Omega$  denote the number of vertices  $v \in T$  that have outdegree  $d_T^+(v) \in \Omega$ . For any value  $n$  that the number  $|T|_\Omega$  can have with positive probability, we let  $T_n^\Omega$  denote the result of conditioning the tree  $T$  on  $|T|_\Omega = n$ . The goal is to describe the limit of  $R^{[t]}(T_n^\Omega, \vec{v})$ , where  $\vec{v} = (v_1, \dots, v_k)$  are independently and uniformly chosen vertices of  $T_n^\Omega$ , *conditioned* to have outdegree in  $\Omega$ .

We first recall the definition of simply and doubly size-biased versions of  $\xi$ , namely the random variables  $\hat{\xi}$  and  $\xi^*$  with distributions

$$\mathbb{P}(\hat{\xi} = i) = i\mathbb{P}(\xi = i), \quad (4.2)$$

$$\mathbb{P}(\xi^* = i) = i(i-1)\mathbb{P}(\xi = i)/\sigma^2. \quad (4.3)$$

Furthermore, for any fixed integer  $k \geq 1$  we say a *proper  $k$ -tree* is a (rooted) plane tree that has precisely  $k$  leaves, labelled from 1 to  $k$ , such that the root has outdegree 1 and all other internal vertices have outdegree 2. Note that each such tree has  $2k - 1$  edges and that there are  $k! \text{Cat}_{k-1} = 2^{k-1} \prod_{i=1}^{k-1} (2i - 1)$  such trees. Indeed, up to the single edge attached to the root, these trees are complete binary trees with  $k$  leaves and a labelling of these leaves. In the following, we order the edges of proper  $k$ -tree in some canonical order (e.g. depth first search order), so that we can speak of the  $i$ -th edge of the tree; the chosen order is not relevant though.

For each integer  $t \geq 1$  we can now construct a random rooted plane tree  $T_\Omega^{k,t}$  with  $k$  distinguished vertices labelled from 1 to  $k$  having outdegree in  $\Omega$ , and  $2k - 1$  edges having length-labels. We will prove later that this tree is the limit of  $R^{[t]}(T_n^\Omega, \vec{v})$ . A special case of this construction is illustrated in Fig. 8. The general procedure goes as follows:

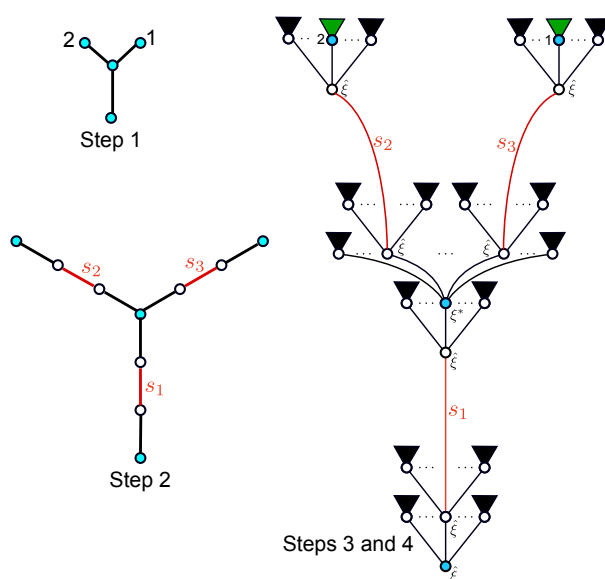


Figure 8: The construction of the limit tree  $T_{\Omega}^{k,t}$  for  $k = 2$  and  $t = 1$ . The essential vertices are coloured blue, and the middle edges are coloured red. Each occurrence of  $\hat{\xi}$  or  $\xi^*$  at the side of a vertex represents that this vertex receives offspring according to an independent copy of the corresponding random variable (step 3). Each black triangle represents an independent copy of the Galton–Watson tree  $T$  (step 4). The green triangles represent independent copies of  $T$  conditioned on having root degree in  $\Omega$  (step 4).

1. (*Pick a skeleton*) Draw a proper  $k$ -tree uniformly at random. Its leaves will correspond to the distinguished labelled vertices of  $T_{\Omega}^{k,t}$ . Each possible outcome of this step is attained with probability

$$\frac{1}{2^{k-1} \prod_{i=1}^{k-1} (2i-1)}.$$

2. (*Stretch it*) Select a vector  $\vec{s} = (s_i)_i \in \mathbb{R}_{>0}^{2k-1}$  at random with density

$$(3 \cdot 5 \cdots (2k-3)) (\sum_i s_i) \exp \left( -\frac{(\sum_i s_i)^2}{2} \right). \quad (4.4)$$

It is easy to check that this defines a probability distribution, using classical expressions for absolute moments of Gaussian distribution. For each  $1 \leq i \leq 2k-1$ , we replace the  $i$ -th edge of the  $k$ -tree by a path of length  $2t+1$  and assign label  $s_i$  to the central edge of this path.

3. (*Thicken it*) Each internal vertex receives additional offspring, independently from the rest. Here vertices with outdegree 1 receive additional offspring according to an independent copy of  $\hat{\xi} - 1$ , while vertices with outdegree 2 receive additional offspring according to an independent copy of  $\xi^* - 2$ . An ordering of the total offspring that respects the ordering of the pre-existing offspring is chosen uniformly at random.
4. (*Graft branches*) Each distinguished vertex (i.e., each leaf of the original  $k$ -tree) becomes the root of an independent copy of  $T$  conditioned on having root-degree in  $\Omega$ . Other leaves of the tree resulting from step 3 become the roots of independent copies of Galton–Watson trees  $T$ , without conditioning.



**Lemma 4.1.** *Seen as an element in  $\mathcal{T}_{k,\Omega}^{[t]} \times (\mathbb{R}_+)^{2k-1}$ , the random tree  $T_\Omega^{k,t}$  has density*

$$f(G, \vec{u}) = p_G h(\vec{u}) \mathbb{1}[G \text{ generic}],$$

where, for a generic  $G$  in  $\mathcal{T}_{k,\Omega}^{[t]}$  and  $\vec{u}$  in  $(\mathbb{R}_+)^{2k-1}$ , we have

$$p_G := \frac{\mathbb{P}(\xi \in \Omega)^{-k}}{\sigma^{2k-2} \prod_{i=1}^{k-1} (2i-1)} \prod_{v \in V_G} \mathbb{P}(\xi = d_G^+(v)); \quad (4.5)$$

$$h(\vec{u}) := \left[ \prod_{i=1}^{k-1} (2i-1) \right] (\sum_i u_i) \exp\left(-\frac{(\sum_i u_i)^2}{2}\right). \quad (4.6)$$

*Proof.* By construction, it is clear that  $\text{Sh}(T_\Omega^{k,t})$  and  $\text{Lab}(T_\Omega^{k,t})$  are independent, that  $\text{Sh}(T_\Omega^{k,t})$  is generic and that  $\text{Lab}(T_\Omega^{k,t})$  has density  $h$ . We only need to prove that, for any generic  $G$  in  $\mathcal{T}_{k,\Omega}^{[t]}$ ,  $\mathbb{P}[\text{Sh}(T_\Omega^{k,t}) = G] = p_G$ .

We follow the construction of  $T_\Omega^{k,t}$ . The event  $\text{Sh}(T_\Omega^{k,t}) = G$  holds if and only if the following events occur.

- At step 1, we choose the proper  $k$ -tree corresponding to the genealogy of the distinguished vertices of  $G$ : this happens with probability

$$\frac{1}{2^{k-1} \prod_{i=1}^{k-1} (2i-1)}.$$

- If we choose the correct proper  $k$ -tree, after step 2, the vertices of the resulting tree correspond to the vertices of  $R(G, \vec{u})$ . Then, at step 3, we need to choose for each of them the correct number of children and the correct ordering of these children. For a branching vertex of outdegree  $d$  in  $R(G, \vec{u})$ , the correct number of children is chosen with probability  $\mathbb{P}(\xi^* = d)$  and the correct ordering with probability  $\binom{d}{2}^{-1}$ . Multiplying these probabilities gives

$$\frac{\mathbb{P}(\xi^* = d)}{\binom{d}{2}} = \frac{2}{\sigma^2} \mathbb{P}(\xi = d).$$

For a non-branching internal vertex of outdegree  $d$  in  $R(G, \vec{u})$ , the correct number of children is chosen with probability  $\mathbb{P}(\hat{\xi} = d)$  and the correct ordering with probability  $d^{-1}$ . Again, multiplying these two, we get

$$\frac{\mathbb{P}(\hat{\xi} = d)}{d} = \mathbb{P}(\xi = d).$$

- In step 4 of the construction, we need to choose copies of  $T$  or  $\bar{T}$  conditioned to have root outdegree in  $\Omega$  (the black and green triangles in Fig. 8) corresponding to that in  $G$ . The probability of this event is given as a product as follows. For each distinguished vertex  $v$  of outdegree  $d$ , we have a factor  $\mathbb{P}(\xi = d)/\mathbb{P}(\xi \in \Omega)$  (the denominator comes from the conditioning that the outdegree of such vertex is in  $\Omega$ ). For vertices in  $G \setminus R(G, \vec{u})$  of outdegree  $d$ , we have a factor  $\mathbb{P}(\xi = d)$ .

Summing up, since there are  $k-1$  branching vertices and  $k$  distinguished ones, we get that

$$\mathbb{P}[\text{Sh}(T_\Omega^{k,t}) = G] = \frac{1}{2^{k-1} \prod_{i=1}^{k-1} (2i-1)} \frac{2^{k-1}}{\sigma^{2k-2}} \mathbb{P}(\xi \in \Omega)^{-k} \prod_{v \in V_G} \mathbb{P}(\xi = d_G^+(v)), \quad (4.7)$$

and the factors  $2^{k-1}$  in the numerator and denominator cancel out.  $\square$

#### 4.4 Convergence

The following lemma extends Aldous' skeleton decomposition [6, Eq. (49)] by keeping track of  $o(\sqrt{n})$ -neighbourhoods near the essential vertices of the skeleton. The  $o(\sqrt{n})$ -threshold is sharp (for the applications in this paper, the convergence of  $t_n$ -neighbourhoods for any sequence  $t_n$  tending to infinity would suffice). We note that  $o(\sqrt{n})$ -neighbourhoods of the root have been previously considered in the literature, e.g. by Aldous [4, 5] and Kersting [36]; see also [51, Theorem 5.2] for a result on the  $o(\sqrt{n})$ -neighbourhood of a uniform random vertex in the tree. Besides, Lemma 4.2 is also related to scaling limits obtained by Kortchemski [37] and Rizzolo [46], that imply convergence of  $R(\mathbf{T}_n^\Omega, \vec{\mathbf{v}})$ .

We recall that we see trees of the form  $s.R^{[t]}(T, \vec{u})$  and  $\mathbf{T}_\Omega^{k,t}$  as elements of the space  $\mathcal{T}_{k,\Omega}^{[t]} \times (\mathbb{R}_+)^{2k-1}$  as explained in Section 4.2.

**Lemma 4.2.** *Suppose that the offspring distribution  $\xi$  is critical, aperiodic, and has finite variance  $\sigma^2$ . Let  $\vec{\mathbf{v}}$  be a vector of  $k \geq 1$  independently and uniformly selected vertices with outdegree in  $\Omega$  of the conditioned tree  $\mathbf{T}_n^\Omega$ . Then for each constant positive integer  $t$  it holds that*

$$c_\Omega \sigma n^{-1/2} . R^{[t]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}) \xrightarrow{d} \mathbf{T}_\Omega^{k,t} \quad (4.8)$$

with  $c_\Omega = \sqrt{\mathbb{P}(\xi \in \Omega)}$ . Even stronger, for each sequence  $t_n = o(\sqrt{n})$  of positive integers it holds that

$$\sup_{A,B} \left| \mathbb{P}[c_\Omega \sigma n^{-1/2} . R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}) \in A \times B] - \mathbb{P}[\mathbf{T}_\Omega^{k,t_n} \in A \times B] \right| \rightarrow 0, \quad (4.9)$$

with  $A$  ranging over all subsets of  $\mathcal{T}_{k,\Omega}^{[t_n]}$ , and  $B$  over open intervals of  $(\mathbb{R}_+)^{2k-1}$ .

*Proof.* We fix some sequence  $t_n$  and let, for each  $n$ ,  $(G, \vec{x})$  be an element in  $\mathcal{T}_{k,\Omega}^{[t_n]} \times (\mathbb{R}_+)^{2k-1}$ , with  $G$  generic and  $\vec{x}$  taking integer coordinates.

We also fix constants  $b > a > 0$  and a sequence  $s_n$  with  $s_n = o(n)$ . The core of the proof consists in establishing that, as  $n \rightarrow \infty$ , we have

$$\mathbb{P}[1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}) = (G, \vec{x})] \sim \left( \frac{\sigma c_\Omega}{\sqrt{n}} \right)^{2k-1} h\left(\sigma c_\Omega n^{-1/2} \vec{x}\right) p_G, \quad (4.10)$$

uniformly on pairs  $(G, \vec{x})$  such that  $\sum_{i=1}^{2k-1} x_i$  is in  $[a\sqrt{n}, b\sqrt{n}]$  and  $|G|_\Omega \leq s_n$  (recall that  $p_G$  is defined in (4.5) and  $h(\cdot)$  in (4.6)).

Assume temporarily (4.10). Summing over all possible values of  $\vec{x}$  (such that  $\sum_{i=1}^{2k-1} x_i$  lies in  $[a\sqrt{n}, b\sqrt{n}]$ , and making  $a$  go to 0,  $b$  go to  $+\infty$ ), we have

$$\mathbb{P}[\text{Sh}(1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}})) = G] = \mathbb{P}[\text{Sh}(c_\Omega \sigma n^{-1/2} . R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}})) = G] \sim p_G, \quad (4.11)$$

uniformly on trees  $G$  with  $|G|_\Omega \leq s_n$ . Moreover, conditionally on the shape of this skeleton being  $G$ , Eq. (4.10) gives a local limit theorem for  $\text{Lab}(1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}))$  with scaling factor  $c_\Omega \sigma n^{-1/2}$  and limiting distribution with density  $h$ . This implies convergence in distribution of  $\text{Lab}(c_\Omega \sigma n^{-1/2} . R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}))$  to a random variable of density  $h$ . Comparing with Theorem 4.1, we see that Eq. (4.10) implies Eq. (4.8). Proving Eq. (4.9) needs an extra ingredient and we come back to it at the end of the proof.

To prove Eq. (4.10), we need some additional notation. First, we write  $\ell = \sum_{i=1}^{2k-1} x_i$ . Additionally, we let  $(\mathbf{X}_i, \hat{\xi}_i)_{i \geq 1}$  be independent copies of  $|T|_\Omega$  and  $\hat{\xi}$ . Finally, we also set

$$\mathbf{S}_m := \mathbf{X}_1 + \dots + \mathbf{X}_m \text{ (for } m \geq 0), \quad \mathbf{Q} := \sum_{i=1}^{\ell} \mathbf{1}_{\hat{\xi}_i \in \Omega} \text{ and } \mathbf{L} := \sum_{i=1}^{\ell} (\hat{\xi}_i - 1). \quad (4.12)$$

The proof of Eq. (4.10) is splitted in two parts, respectively of combinatorial and analytic nature. The combinatorial part shows that

$$\begin{aligned} \mathbb{P}[1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}) = (G, \vec{x})] &= n^{-k} \left( \prod_{v \in V_G} \mathbb{P}(\xi = d_G^+(v)) \right) \frac{\mathbb{P}(\mathbf{S}_L = n - |G|_\Omega - \mathbf{Q})}{\mathbb{P}(|\mathbf{T}|_\Omega = n)} \\ &= n^{-k} c_\Omega^{2k} \sigma^{2k-2} \left( \prod_{i=1}^{k-1} (2i-1) \right) p_G \frac{\mathbb{P}(\mathbf{S}_L = n - |G|_\Omega - \mathbf{Q})}{\mathbb{P}(|\mathbf{T}|_\Omega = n)}. \end{aligned} \quad (4.13)$$

We do this by decomposing combinatorially pairs  $(T_\star, \vec{v}_\star)$  (i.e. trees with distinguished vertices) such that  $1.R^{[t_n]}(T_\star, \vec{v}_\star) = (G, \vec{x})$ .

The analytic part, based on a standard local limit lemma, then analyzes the numerator of the last factor and shows that

$$\mathbb{P}(\mathbf{S}_L = n - |G|_\Omega - \mathbf{Q}) \sim (2\pi)^{-1/2} \sigma c_\Omega \ell n^{-3/2} \exp\left(-\frac{\sigma^2 \mathbb{P}(\xi \in \Omega) \ell^2}{2n}\right) \quad (4.14)$$

uniformly on integers  $\ell$  in  $[a\sqrt{n}, b\sqrt{n}]$ , and on trees  $G$  with  $|G|_\Omega \leq s_n$ .

Finally, an estimate for the denominator in (4.13) is given e.g. in [37, Thm. 8.1]:

$$\mathbb{P}(|\mathbf{T}|_\Omega = n) \sim \frac{c_\Omega}{\sqrt{2\pi}\sigma} n^{-3/2}. \quad (4.15)$$

We leave the reader check that, after many obvious cancellations, plugging in the estimates (4.14) and (4.15) into (4.13) gives indeed (4.10).

*The combinatorial part: proof of (4.13).* We first consider the unconditioned Galton-Watson tree  $T$ , and conditionally on  $T$ , a uniform list  $\vec{\mathbf{v}}$  of  $k$  vertices with outdegree in  $\Omega$  in  $T$  (possibly with repetitions). A pair  $(T_\star, \vec{v}_\star)$ , where  $\vec{v}_\star$  is a list of  $k$  vertices with outdegree in  $\Omega$  in the tree  $T_\star$ , is called *good* if  $|T_\star|_\Omega = n$  and  $1.R^{[t_n]}(T_\star, \vec{v}_\star) = (G, \vec{x})$ . Then we have

$$\begin{aligned} \mathbb{P}[1.R^{[t_n]}(\mathbf{T}, \vec{\mathbf{v}}) = (G, \vec{x}), |\mathbf{T}|_\Omega = n] &= \sum_{(T_\star, \vec{v}_\star) \text{ good}} \mathbb{P}(T = T_\star) \mathbb{P}(\vec{\mathbf{v}} = \vec{v}_\star | T = T_\star) \\ &= n^{-k} \sum_{(T_\star, \vec{v}_\star) \text{ good}} \mathbb{P}(T = T_\star). \end{aligned} \quad (4.16)$$

Good pairs  $(T_\star, \vec{v}_\star)$  can be constructed as follows.

- i) We start from  $(G, \vec{x})$  and replace each edge with a length label  $x_i$  by a path with  $x_i$  internal vertices; in total, this operation creates  $\ell$  new vertices, which we will refer to as the *remote* vertices.
- ii) We choose the outdegrees  $(d_i)_{i \leq \ell}$  in  $T_\star$  of the  $\ell$  remote vertices of  $(T, \vec{u})$ ;
- iii) For each remote vertex, choose the distinguished offspring along which we have to proceed to get to the first descendant that is an essential vertex ( $d_i$  possible choices).
- iv) On each of the  $m := \sum d_i - 1$  other children of the remote vertices, we attach a fringe subtree tree  $(A_j)_{j \leq m}$ .
- v) To ensure that  $|T_\star|_\Omega = n$ , the degrees  $(d_i)_{i \leq \ell}$  and the subtrees  $(A_j)_{j \leq m}$  should be chosen such that  $|G|_\Omega + \sum_{i=1}^\ell \mathbb{1}_{d_i \in \Omega} + \sum_{j=1}^m |A_j|_\Omega = n$ .

Moreover, if  $(T_*, \vec{v}_*)$  corresponds in this construction to sequences  $(d_i)_{i \leq \ell}$  and  $(A_j)_{j \leq m}$ , then we have  $(V_G$  denoting the set of vertices of the tree  $G)$

$$\mathbb{P}(T = T_*) = \left( \prod_{v \in V_G} \mathbb{P}[\xi = d_G^+(v)] \right) \left( \prod_{i=1}^{\ell} \mathbb{P}[\xi = d_i] \right) \left( \prod_{j=1}^m \mathbb{P}[T = A_j] \right).$$

(This probability is independent from the choices made in step iii)). The sum over good pairs  $(T_*, \vec{v}_*)$  in Eq. (4.16) can be rewritten as a sum over sequences of positive integers  $(d_i)_{i \leq \ell}$  and sequences of trees  $(A_j)_{j \leq m}$ , with an extra factor  $\prod_i d_i$  coming from the choices in item iii) above. We get

$$\begin{aligned} \mathbb{P}[1.R^{[t_n]}(T, \vec{v}) = (G, \vec{x}), |T|_{\Omega} = n] \\ = n^{-k} \left( \prod_{v \in V_G} \mathbb{P}[\xi = d_G^+(v)] \right) \sum_{d_1, \dots, d_{\ell} \geq 1} \left[ \prod_{i=1}^{\ell} d_i \mathbb{P}[\xi = d_i] \right. \\ \left. \cdot \sum_{A_1, \dots, A_m \text{ trees}} \mathbb{1} \left[ \sum_{j=1}^m |A_j|_{\Omega} = n - |G|_{\Omega} - \sum_{i=1}^{\ell} \mathbb{1}_{d_i \in \Omega} \right] \prod_{j=1}^m \mathbb{P}[T = A_j] \right]. \end{aligned}$$

The sum in the last line is the probability that the total number of vertices with outdegree in  $\Omega$  in  $m$  independent copies of  $T$  is  $n - |G|_{\Omega} - \sum_{i=1}^{\ell} \mathbb{1}_{d_i \in \Omega}$ , i.e., with the notation Eq. (4.12), this is  $\mathbb{P}(S_m = n - |G|_{\Omega} - \sum_{i=1}^{\ell} \mathbb{1}_{d_i \in \Omega})$ . Recalling that by definition,  $m = \sum_i d_i - 1$ , we get

$$\begin{aligned} \mathbb{P}[1.R^{[t_n]}(T, \vec{v}) = (G, \vec{x}), |T|_{\Omega} = n] = n^{-k} \left( \prod_{v \in V_G} \mathbb{P}[\xi = d_G^+(v)] \right) \cdot \\ \left[ \sum_{d_1, \dots, d_{\ell} \geq 1} \left( \prod_{i=1}^{\ell} d_i \mathbb{P}[\xi = d_i] \right) \mathbb{P} \left( S_{\sum_{i=1}^{\ell} (d_i - 1)} = n - |G|_{\Omega} - \sum_{i=1}^{\ell} \mathbb{1}_{d_i \in \Omega} \right) \right]. \end{aligned}$$

With the notation Eq. (4.12), the right-hand side can be simplified as

$$\mathbb{P}[1.R^{[t_n]}(T, \vec{v}) = (G, \vec{x}), |T|_{\Omega} = n] = n^{-k} \left( \prod_{v \in V_G} \mathbb{P}[\xi = d_G^+(v)] \right) \mathbb{P}(S_L = n - |G|_{\Omega} - Q).$$

Dividing by  $\mathbb{P}(|T|_{\Omega} = n)$  gives Eq. (4.13), as wanted.

*The analytic part: proof of (4.14).* We are now looking for an asymptotic estimates for the probability  $\mathbb{P}(S_L = n - |G|_{\Omega} - Q)$ . Since  $S_m$  is a sum of  $m$  i.i.d. random variables with the same law as  $|T|_{\Omega}$ , this asymptotics depends on the tail of the distribution of  $|T|_{\Omega}$ . We recall from (4.15) that

$$\mathbb{P}(|T|_{\Omega} = n) \sim (2\pi\sigma_{\Omega}^2)^{-1/2} n^{-3/2},$$

with  $\sigma_{\Omega}^2 = \sigma^2 / \mathbb{P}(\xi \in \Omega) = \sigma^2 / c_{\Omega}^2$ . This implies that  $|T|_{\Omega}$  lies in the domain of attraction of the positive  $(1/2)$ -stable law. Hence by [23, Sec. 50],

$$\lim_{m \rightarrow \infty} \sup_{r \geq 0} |m^2 \mathbb{P}(S_m = r) - \sigma_{\Omega}^2 g(\sigma_{\Omega}^2 r / m^2)| = 0, \quad (4.17)$$

with  $g$  denoting the positive  $(1/2)$ -stable density given by

$$g(x) = (2\pi)^{-1/2} x^{-3/2} \exp(-1/(2x)). \quad (4.18)$$

In particular, since this density is bounded, we have that, for  $m$  large enough,

$$\sup_{r \geq 0} |\mathbb{P}(\mathbf{S}_m = r)| = O(m^{-2}). \quad (4.19)$$

The law of large numbers tells us that

$$\ell^{-1} \mathbf{L} \xrightarrow{\mathbb{P}} \mathbb{E}[\hat{\xi}] - 1 = \mathbb{E}[\xi^2] - 1 = \sigma^2.$$

Moreover, from standard deviation estimates, there is a sequence  $\epsilon_n \rightarrow 0$  with

$$\mathbb{P}(\mathbf{L} \notin (1 \pm \epsilon_n)\sigma^2\ell) \rightarrow 0,$$

and this sequence can be chosen uniformly for all  $\ell$  in  $[a\sqrt{n}, b\sqrt{n}]$ . It follows by conditioning on  $\mathbf{L}$  and using (4.19) that

$$\begin{aligned} \mathbb{P}[\mathbf{S}_\mathbf{L} = n - |G|_\Omega - \mathbf{Q}, \mathbf{L} \in (\sigma^2\ell/2, (1 - \epsilon_n)\sigma^2\ell) \text{ or } \mathbf{L} > (1 + \epsilon_n)\sigma^2\ell] \\ = O(\ell^{-2})\mathbb{P}(\mathbf{L} \notin (1 \pm \epsilon_n)\sigma^2\ell) = o(\ell^{-2}) = o(n^{-1}). \end{aligned}$$

Moreover, the Azuma–Hoeffding inequality implies that for large enough  $M > 0$

$$\mathbb{P}(\mathbf{L} < \sigma^2\ell/2) \leq \mathbb{P}\left(\sum_{i=1}^{\ell} (\hat{\xi}_i \mathbf{1}_{\hat{\xi}_i \leq M} - 1) \leq \sigma^2\ell/2\right) \leq \exp(-\Theta(\ell)) = \exp(-\Theta(\sqrt{n})).$$

Thus we obtain

$$\mathbb{P}(\mathbf{S}_\mathbf{L} = n - |G|_\Omega - \mathbf{Q}) = o(n^{-1}) + \mathbb{P}(\mathbf{S}_\mathbf{L} = n - |G|_\Omega - \mathbf{Q}, \mathbf{L} \in (1 \pm \epsilon_n)\sigma^2\ell).$$

By definition, we have  $\mathbf{Q} \leq \ell$  a.s., so that  $n - |G|_\Omega - \mathbf{Q} = n - o(1)$ , uniformly when  $|G|_\Omega \leq s_n$  and  $\ell$  in  $[a\sqrt{n}, b\sqrt{n}]$ . Using (4.17), we have

$$\mathbb{P}(\mathbf{S}_\mathbf{L} = n - |G|_\Omega - \mathbf{Q}) \sim (2\pi)^{-1/2} \sigma_\Omega^{-1} \sigma^2 \ell n^{-3/2} \exp(-\ell^2 \sigma^4 / (2n\sigma_\Omega^2));$$

indeed, the right-hand side being of order  $\Theta(n^{-1})$  (for  $\ell$  in  $[a\sqrt{n}, b\sqrt{n}]$ ), the error term  $o(n^{-1})$  above can be forgotten. This verifies Eq. (4.14).

*Proof of the statement with  $t_n = o(\sqrt{n})$  (Eq. (4.9)).* Since the equivalent in Eq. (4.10) is uniform on trees  $G$  in  $\mathcal{T}_{k,\Omega}^{[t]}$  with  $|G|_\Omega \leq s_n$ , Eq. (4.10) implies a weaker version of Eq. (4.9), where we let  $A$  range only over subsets of  $\mathcal{T}_{k,\Omega}^{[t]}$  containing only trees with  $\Omega$ -size at most  $s_n$ . It remains to verify that there exists a sequence  $s_n = o(n)$  that additionally satisfies

$$\mathbb{P}[|\text{Sh}(\mathbf{T}_\Omega^{k,t_n})|_\Omega \geq s_n] \rightarrow 0. \quad (4.20)$$

Since by (4.11), we have

$$\mathbb{P}[|\text{Sh}(\mathbf{T}_\Omega^{k,t_n})|_\Omega \leq s_n] - \mathbb{P}[|\text{Sh}(c_\Omega n^{-1/2} \cdot R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}))|_\Omega \leq s_n] \rightarrow 0,$$

(4.20) would imply

$$\mathbb{P}(|\text{Sh}(c_\Omega n^{-1/2} \cdot R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{\mathbf{v}}))|_\Omega \geq s_n) \rightarrow 0$$

and hence complete the proof of Eq. (4.9).

Let us check (4.20). By construction of  $\mathbf{T}_\Omega^{k,t_n}$  (see Fig. 8), we have:

$$|\text{Sh}(\mathbf{T}_\Omega^{k,t_n})|_\Omega \leq 4kt_n + \mathbf{S}_\mathbf{M} + \mathbf{S}'_k, \quad (4.21)$$

where the summands in the right-hand side are independent and distributed as follows:

- $\mathbf{M}$  satisfies

$$\mathbf{M} \stackrel{d}{=} \sum_{i=1}^{k-1} (\xi_i^* - 2) + \sum_{i=1}^{1+t_n(4k-2)} (\hat{\xi}_i - 1),$$

and, as above,  $\mathbf{S}_M$  denotes the sum of  $M$  independent random variables  $(X_i)_{i \leq M}$  of law  $|\mathbf{T}|_\Omega$ , the  $X_i$  being also independent of  $\mathbf{M}$ ;

- $\mathbf{S}'_k$  is the sum of  $k$  i.i.d. random variables of law  $|\mathbf{T}|_\Omega$ , conditioned on the root of  $\mathbf{T}$  having outdegree in  $\Omega$ ;
- $4kt_n$  is an upper bound for the number of vertices on the stretched skeleton of  $\mathbf{T}_\Omega^{k,t_n}$  having outdegree in  $\Omega$ .

It is known (see Rizzolo [46, Thm. 6]) that  $|\mathbf{T}|_\Omega \in \{0, 1, \dots\}$  may be stochastically bounded by the number of vertices of Galton–Watson tree with a different offspring distribution that is also critical and has finite variance. It follows by a general result for the size of Galton–Watson forests, there is a constant  $C > 0$  such that

$$\mathbb{P}(\mathbf{S}_m \geq x) \leq Cmx^{-1/2}. \quad (4.22)$$

for all  $m$  and  $x$ . See Devroye and Janson [19, Lem. 4.3] and Janson [30, Lem. 2.1]. Let  $\epsilon > 0$  be given. We choose a constant  $K > 0$  such that

$$\mathbb{P}[\mathbf{S}_{\sum_{i=1}^{k-1} (\xi_i^* - 2)} + \mathbf{S}'_k > K] \leq \epsilon/2.$$

Letting  $\mathbf{M}' \stackrel{d}{=} \sum_{i=1}^{1+t_n(4k-2)} (\hat{\xi}_i - 1)$  be independent from the family  $(\mathbf{X}_i)_{i \geq 1}$ , it follows by Inequalities (4.21) and (4.22) that

$$\begin{aligned} \mathbb{P}[\text{Sh}(\mathbf{T}_\Omega^{k,t_n})|_\Omega \geq s_n] &\leq \mathbb{P}[4kt_n + \mathbf{S}_M + \mathbf{S}'_k \geq s_n] \leq \epsilon/2 + \mathbb{P}[\mathbf{S}_{M'} \geq s_n - K - 4kt_n] \\ &\leq \epsilon/2 + C(1 + t_n(4k - 2))(\mathbb{E}[\hat{\xi}] - 1)(s_n - K - 4kt_n)^{-1/2}. \end{aligned}$$

Hence (4.20) holds if we select  $s_n = o(n)$  such that  $t_n/\sqrt{s_n} \rightarrow 0$ , which is clearly possible since  $t_n = o(\sqrt{n})$ . This completes the proof.  $\square$

The above proof essentially also gives a local version of Lemma 4.2, which we believe to be of independent interest, and state below as Theorem 4.3.

**Lemma 4.3.** *Let the offspring distribution  $\xi$  be critical, aperiodic, and have a finite variance. Let  $\vec{v}$  be a vector of  $k \geq 1$  independently and uniformly selected vertices with outdegree in  $\Omega$  of the conditioned tree  $\mathbf{T}_n^\Omega$ . Besides, we fix sequences  $t_n$  and  $s_n$  of non-negative integers satisfying  $t_n^2 = o(s_n)$  and  $s_n = o(n)$ . Then there exists sequence  $a_n$  and  $b_n$  tending to 0 and  $+\infty$  respectively such that:*

i) *the estimates*

$$\mathbb{P}[1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{v}) = (G, \vec{x})] \sim \left( \frac{\sigma c_\Omega}{\sqrt{n}} \right)^{2k-1} h\left(\sigma c_\Omega n^{-1/2} \vec{x}\right) p_G, \quad (4.23)$$

*holds uniformly for all pairs  $(G, \vec{x})$  in  $\mathcal{T}_{k,\Omega}^{[t_n]} \times (R_+)^{2k-1}$  with  $G$  generic verifying the conditions  $|G|_\Omega \leq s_n$  and  $a_n \sqrt{n} \leq \|\vec{x}\|_1 \leq b_n \sqrt{n}$ ;*

ii) *and, if we write*

$$(G, \vec{x}) := 1.R^{[t_n]}(\mathbf{T}_n^\Omega, \vec{v}),$$

*then the following properties hold with high probability as  $n$  becomes large:  $G$  is generic,  $|G|_\Omega \leq s_n$  and  $a_n \sqrt{n} \leq \|\vec{x}\|_1 \leq b_n \sqrt{n}$ .*

*Proof.* The estimates i) with a fixed  $a$  instead of  $a_n$  and a fixed  $b$  instead of  $b_n$  has been proved in Eq. (4.10) above. The existence of sequences  $a_n$  and  $b_n$  such that i) holds is a direct consequence, using the following elementary analysis lemma.

Let  $F(A, n)$  be a bivariate function, nonincreasing in  $A$ . We assume that for any  $A > 0$ , we have  $\lim_{n \rightarrow \infty} F(A, n) = 0$ . Then, there exists a sequence  $A_n$  tending to 0 such that  $F(A_n, n)$  tends to 0.

Finally ii) holds for any sequences  $a_n$  and  $b_n$  and any  $s_n$  with  $t_n^2 = o(s_n)$ , as a consequence of Eqs. (4.9) and (4.20).  $\square$

Finally, the following statement will be useful (with  $\Omega = \{0\}$ , i.e. marking leaves) in the special case of separable permutations. It can either be proved as a corollary of Theorem 4.3, or similarly to Lemma 6.2 in [9].

**Corollary 4.4.** *Let the offspring distribution  $\xi$  be critical, aperiodic, and have a finite variance. Let  $\vec{v}$  be a vector of  $k \geq 1$  independently and uniformly selected vertices with outdegree in  $\Omega$  of the conditioned tree  $T_n^\Omega$ . Then, for any fixed  $t$ , asymptotically as  $n \rightarrow \infty$ , the parities of the heights of the essential vertices induced by  $\vec{v}$  (except the root of  $T_n^\Omega$ ) converge to Bernoulli random variables of parameter  $1/2$ , independent among themselves, and from the tree  $\text{Sh}(1.R^{[t]}(T_n^\Omega, \vec{v}))$ .*

## 5 Scaling limits

### 5.1 Background on permuton convergence

As said in introduction, a *permuton*  $\mu$  is a Borel probability measure on the unit square  $[0, 1]^2$  with uniform marginals, that is

$$\mu([0, 1] \times [a, b]) = \mu([a, b] \times [0, 1]) = b - a$$

for all  $0 \leq a \leq b \leq 1$ . Any permutation  $\nu$  of size  $n \geq 1$  may be interpreted as a permuton  $\mu_\nu$  given by the linear combination of area measures

$$\mu_\nu = n \sum_{i=1}^n \delta_{[(i-1)/n, i/n] \times [(\nu(i)-1)/n, \nu(i)/n]}.$$

By definition, a random permutation  $\nu_n$  converges weakly to a random permuton  $\mu$  as  $n \rightarrow \infty$  if the random probability measure  $\mu_{\nu_n}$  converges weakly to  $\mu$ . There are different characterisations for this form of convergence [7, Thm. 2.5]. In particular, if  $\nu_n$  has size  $n$ , then the following statements are equivalent:

- i) There exists a permuton  $\mu$  such that  $\mu_{\nu_n} \xrightarrow{d} \mu$ .
- ii) For any integer  $k \geq 1$  the pattern  $\text{pat}_{\mathbf{I}_{n,k}}(\nu_n)$  induced by a uniform random  $k$ -element subset  $\mathbf{I}_{n,k} \subseteq [n]$  admits a distributional limit  $\rho_k$ .

In this case, the limit family  $(\rho_k)_k$  is consistent in the sense that  $\rho_k$  has size  $k$  a.s. for all  $k$  and  $\text{pat}_{\mathbf{I}_{n,k}}(\rho_n) \stackrel{d}{=} \rho_k$  for all  $1 \leq k \leq n$ . The permuton  $\mu$  may be constructed from the family  $(\rho_k)_{k \geq 1}$ , and is hence uniquely determined by it. In fact, there is a bijection between random permutons and consistent families [7, Prop. 2.9]. (Compare with a similar result for random trees [6, Thm. 18].)

The following permutons were introduced in [7, 9] where they were proved to be the limit of some substitution-closed classes. (See also [40] for some properties of these permutons.)

- i) The Brownian separable permuton corresponds to the case where  $\rho_k$  is the image by  $\text{CT}^{-1}$  of a uniform binary plane tree with  $k$  leaves with uniform independent decorations from  $\{\oplus, \ominus\}$  on its internal vertices. (Recall from Theorem 2.9 that  $\text{CT}^{-1}$  can be applied to  $\{\oplus, \ominus\}$ -decorated trees, where neighbours may have the same sign.)
- ii) Let  $0 < p < 1$  be a constant. The biased Brownian separable permuton of parameter  $p$  is constructed in the same way, but instead of assigning the  $\oplus / \ominus$  decorations via fair coin flips, we toss a biased coin that shows  $\oplus$  with probability  $p$ .

Putting together the pattern characterization of permuton convergence (recalled above), this description of  $\rho_k$ , and the connection between patterns and subtrees explained in Section 2.5, we get a convenient sufficient condition for the convergence to a (biased) separable Brownian permuton.

To state it, we recall that, if  $\vec{\ell}$  is an ordered sequence of marked leaves in a tree  $T$ , then  $R(T, \vec{\ell})$  denotes the subtree consisting of these marked leaves and all their ancestors. In addition, we denote by  $R^*(T, \vec{\ell})$  the tree obtained from  $R(T, \vec{\ell})$  by successively removing all non-root vertices of outdegree 1, merging their two adjacent edges.

**Lemma 5.1.** *Let  $p$  be a constant in  $[0, 1]$  and, for each  $n \geq 1$ ,  $\nu_n$  be a random permutation of size  $n$ . For each fixed  $k \geq 1$ , we take a uniform random sequence  $\vec{\ell} = (\ell_1, \dots, \ell_k)$  of  $k$  leaves in the canonical tree  $T_n$  of  $\nu_n$ . We make the following assumptions.*

- *The tree  $R^*(T_n, \vec{\ell})$  should converge (in distribution) to a proper  $k$ -tree.*
- *For each non-root internal vertex  $u$  of  $R^*(T_n, \vec{\ell})$ , we choose arbitrarily two leaves from  $\vec{\ell}$ , say  $\ell_{i_u}$  and  $\ell_{j_u}$ , whose common ancestor is  $u$ . We then assume that  $\ell_{i_u}$  and  $\ell_{j_u}$  form a non-inversion asymptotically with probability  $p$ , and that, when  $u$  runs over non-root internal vertices of  $R^*(T_n, \vec{\ell})$ , these events are asymptotically independent from each other and from the shape  $R^*(T_n, \vec{\ell})$ .*

Then  $\nu_n$  converges to the biased separable Brownian permuton of parameter  $p$ .

The arbitrary choices made in the second item above are irrelevant. Indeed, when  $u$  has out-degree 2 in  $R^*(T_n, \vec{\ell})$  (which is the case with high probability under the first assumption), the fact that  $\ell_{i_u}$  and  $\ell_{j_u}$  form an inversion or not does not depend on the choice of  $\ell_{i_u}$  and  $\ell_{j_u}$  (this an easy consequence of the discussion from Section 2.5).

## 5.2 Permuton convergence of random permutations from substitution-closed classes

We now prove our first main theorem, Theorem 1.1. We start by stating this theorem more precisely.

**Theorem 5.2.** *Let  $\nu_n$  be the uniform  $n$ -sized permutation from a proper substitution-closed class of permutations  $\mathcal{C}$ . Let  $\xi$  be the offspring distribution of the associated Galton–Watson tree model. Suppose that  $\mathbb{E}[\xi] = 1$  and  $\mathbb{V}[\xi] < \infty$ . That is, either*

$$S'(\rho_S) > \frac{2}{(1 + \rho_S)^2} - 1,$$

or

$$S'(\rho_S) = \frac{2}{(1 + \rho_S)^2} - 1 \quad \text{and} \quad S''(\rho_S) < \infty.$$

Then

$$\mu_{\nu_n} \xrightarrow{d} \mu^{(p)},$$



with  $\mu^{(p)}$  denoting the biased Brownian separable permuton with parameter

$$p = \frac{2}{\sigma^2} (\kappa(1 + \kappa)^3 \text{Occ}_{12}(\kappa) + \kappa), \quad (5.1)$$

where  $\kappa$  and  $\sigma^2 = \mathbb{V}[\xi]$  are defined in Theorem 3.5 and  $\text{Occ}_{12}(z) = \sum_{\alpha \in \mathcal{S}} \text{occ}(12, \alpha) z^{|\alpha|-2}$ , with  $\text{occ}(12, \alpha)$  being the number of occurrences of the pattern 12 in  $\alpha$ .

This includes the case where  $\mathcal{C}$  is the class of separable permutations, for which  $\mathcal{S}(z) = 0$  and  $p = 1/2$ .

*Proof.* By Proposition 3.2, it suffices to show that the uniform  $n$ -sized permutation  $\nu_n$  from  $\mathcal{C}_{\text{not}\oplus}$  satisfies

$$\mu_{\nu_n} \xrightarrow{d} \mu^{(p)}.$$

We first consider the separable case  $\mathcal{S} = \emptyset$ . Let  $T_n$  be the canonical tree of  $\nu_n$ . Here a vertex of  $T_n$  is decorated with  $\ominus$  if and only if it has even height. Without its decorations,  $T_n$  has the law of a critical Galton–Watson tree with finite variance conditioned on having  $n$  leaves (see [9, Sec. 2.2] or Section 3; for the separable case, packed trees and canonical trees only differ by their decorations).

Let  $k \geq 1$  be given and  $\vec{\ell} = (\ell_1, \dots, \ell_k)$  be a uniform random sequence of leaves in  $T_n$ . It follows from Lemma 4.2 that  $R^*(T_n, \vec{\ell})$  is asymptotically a uniform random proper  $k$ -tree. Corollary 4.4 yields the additional information that the parities of the lengths of the  $2k - 1$  paths in  $T_n$  corresponding to the edges of  $R^*(T_n, \vec{\ell})$  converge jointly to  $2k - 1$  independent fair coin flips, independently of the shape  $R^*(T_n, \vec{\ell})$ . Hence in the limit as  $n \rightarrow \infty$  each non-root internal vertex of  $R^*(T_n, \vec{\ell})$  receives a sign  $\oplus$  or  $\ominus$  with probability  $1/2$  (meaning that the corresponding leaves, in the sense of the second item of Theorem 5.1, form an inversion with probability  $1/2$ ); moreover, these events are asymptotically independent from each other and from the shape  $R^*(T_n, \vec{\ell})$ . As this holds for all  $k \geq 1$ , thanks to Theorem 5.1, it follows that  $\nu_n$  converges in distribution to the Brownian separable permuton  $\mu^{(1/2)}$ .

Let us now consider the case  $\mathcal{S} \neq \emptyset$ . In this case, it is more convenient to work with packed trees rather than canonical trees (note however that both trees have the same set of leaves). In particular, the random packed tree  $P_n = (T_n, \lambda_{T_n})$  associated with the uniform permutation  $\nu_n$  in  $\mathcal{C}_{\text{not}\oplus}$  is a Galton–Watson tree with a specific offspring distribution  $\xi$  conditioned on having  $n$  leaves, with independent random decorations on each vertex (see Section 3). As before, we fix  $k \geq 1$  and consider a uniformly selected set of distinct leaves  $\vec{\ell} = (\ell_1, \dots, \ell_k)$  in  $T_n$ .

By Theorem 4.2, we know that  $R^*(T_n, \vec{\ell})$  converges (in distribution) to a uniform proper  $k$ -tree (recall that  $\xi$  is always aperiodic and that it has expectation 1 and finite variance by assumption, as needed to apply Theorem 4.2). In particular, the tree  $R^*(T_n, \vec{\ell})$  is a proper tree (with a root of outdegree 1 and other internal vertices of outdegree 2) with high probability, as  $n \rightarrow \infty$ . When this is the case, since the packing construction only merges internal vertices,  $R^*(T_n, \vec{\ell})$  coincide with  $R^*(\tilde{T}_n, \vec{\ell})$ , where  $\tilde{T}_n$  is the canonical tree associated with  $\nu_n$ . Therefore, although Theorem 5.1 is stated with the canonical tree  $\tilde{T}_n$ , we can use it here with the packed tree  $T_n$  instead.

Using the notation of Theorem 5.1, it remains to analyse whether  $\ell_{i_u}$  and  $\ell_{j_u}$  form an inversion or not (for non-root internal vertices  $u$  of  $R^*(T_n, \vec{\ell})$ ).

We recall (see Section 2.5) that if  $u$  is decorated with an  $\mathcal{S}$ -gadget, then whether  $\ell_{i_u}$  and  $\ell_{j_u}$  form an inversion or not is determined by the decoration of  $u$  and by which branches attached to  $u$  contain  $\ell_{i_u}$  and  $\ell_{j_u}$ . This information is contained in  $\text{Sh}(s, R^{[0]}(T_n, \vec{\ell}))$  for any  $s > 0$ .

On the other hand, if  $u$  is decorated with  $\otimes$ , then in order to determine whether  $\ell_{i_u}$  and  $\ell_{j_u}$  form an inversion or not, we have to recover the parity of the distance of  $u$  to its first ancestor decorated with an  $\mathcal{S}$ -gadget (if it exists, otherwise to the root of  $T_n$ ).

Take  $t_n$  tending to infinity, but with  $t_n = o(\sqrt{n})$ . By Lemma 4.2,  $u$  has asymptotically  $t_n$  ancestors with out-degrees  $\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_{t_n}$  being independent copies of  $\hat{\xi}$  defined in (4.2). Moreover the vertex  $u$  and each of its ancestors receive a decoration that gets drawn independently and uniformly at random among all  $\widehat{\mathcal{G}(\mathcal{S})}$ -decorations with size equal to the out-degree of the vertex. In this setting, with high probability, one of the  $t_n$  ancestors will receive an  $\mathcal{S}$ -gadget as decoration. Therefore, with high probability, whether  $\ell_{i_u}$  and  $\ell_{j_u}$  form an inversion is determined by  $\text{Sh}(s.R^{[t_n]}(T_n, \vec{\ell}))$  for any  $s > 0$ .

We say that two families (indexed by  $\mathbb{N}$ ) of probability distributions are *close* when their total variation distance tends to 0 as  $n$  tends to infinity. By Lemma 4.2, the distributions of the random trees  $\text{Sh}(s_n.R^{[t_n]}(T_n, \vec{\ell}))$  and  $\text{Sh}(T_{\{0\}}^{k, t_n})$  are close, for a well-chosen sequence  $s_n$ . From the above discussion, this implies that the joint distributions of  $R^*(T_n, \vec{\ell})$  and

$$\left( \mathbb{1} \left[ \ell_{i_u} \text{ and } \ell_{j_u} \text{ form an inversion in } (T_n, \vec{\ell}) \right] \right)_u$$

are close to the distributions of the same variables in the limiting tree  $T_{\{0\}}^{k, t_n}$ . When  $n$  tends to infinity, these tend a.s. (with the obvious coupling between the  $T_{\{0\}}^{k, t_n}$ ) to the same variables in  $T_{\{0\}}^{k, t^*}$ , where  $t^*$  denotes the minimal radius such that each internal  $\otimes$ -decorated essential vertex (different from the root) has an ancestor decorated by an  $\mathcal{S}$ -gadget.

In the limiting tree  $T_{\{0\}}^{k, t^*}$ , the neighbourhoods of the essential vertices  $u$  are independent from each other and all have the same distribution (which does not depend on  $k$ , nor on the shape  $R^*(T_{\{0\}}^{k, t^*}, \vec{\ell})$ , the latter being the proper  $k$ -tree taken at step 1 of the construction). Therefore the probability that  $\ell_{i_u}$  and  $\ell_{j_u}$  form a non-inversion tend to some parameter  $p$  in  $[0, 1]$ , which depends only on the permutation class  $\mathcal{C}$  we are working with. Moreover these events are asymptotically independent from each other and from the shape  $R^*(T_n, \vec{\ell})$ . From Theorem 5.1, this implies that  $\mu_{\nu_n} \xrightarrow{d} \mu^{(p)}$ .

It remains to calculate an explicit expression for the limiting probability  $p$ . For this, we consider  $k = 2$ , i.e.,  $p$  is the probability that, in the limiting tree  $T_{\{0\}}^{2, t^*}$ , the two marked leaves  $\ell_1$  and  $\ell_2$  form a non-inversion.

For each integer  $m \geq 1$ , let  $G_m$  be drawn uniformly at random among all  $m$ -sized  $\widehat{\mathcal{G}(\mathcal{S})}$ -objects, i.e.,

$$\mathbb{P}(G_m = G) = 1/q_m, \quad \text{for all } G \in \widehat{\mathcal{G}(\mathcal{S})} \text{ of size } m,$$

where we recall that  $\mathcal{Q}(z) = \widehat{\mathcal{G}(\mathcal{S})}(s) = \sum_{k \geq 2} q_k z^k$  is the generating function in Eq. (3.5). We also recall the following three distributions (see Eqs. (3.12), (4.2) and (4.3))

$$\mathbb{P}(\xi = k) = q_k t_0^{k-1}, \quad \mathbb{P}(\hat{\xi} = k) = k q_k t_0^{k-1}, \quad \mathbb{P}(\xi^* = k) = k q_k (k-1) t_0^{k-1} / \sigma^2, \quad (5.2)$$

where  $t_0 = \frac{\kappa}{1+\kappa}$  is the parameter determined in Theorem 3.5 as the unique number such that

$$\mathcal{S}'(\kappa) = 2/(1+\kappa)^2 - 1, \quad (5.3)$$

$\mathcal{S}(z)$  being the generating functions for simple permutations in the considered substitution-closed class  $\mathcal{C}$ .

To determine whether  $\ell_1$  and  $\ell_2$  form an inversion or not, there are two cases to consider, depending on whether the decoration  $\lambda_{T_{\{0\}}^{k, t^*}}(u) =: \lambda(u)$  of the closest common ancestor  $u$  of  $\ell_1$  and  $\ell_2$  is an  $\mathcal{S}$ -gadget or not.

We start with the case where it is not. Let  $u'$  be the closest ancestor of  $u$  that is decorated with an  $\mathcal{S}$ -gadget. The limiting probability for  $\ell_1$  and  $\ell_2$  to form a non-inversion

in this case is given by

$$\begin{aligned} \mathbb{P}(\lambda(\mathbf{u}) = \otimes, d(\mathbf{u}, \mathbf{u}') \text{ is even and } > 0) \\ = \sum_{k \geq 2} \mathbb{P}(d(\mathbf{u}, \mathbf{u}') \text{ is even} | \lambda(\mathbf{u}) = \otimes) \mathbb{P}(\mathbf{G}_k = \otimes) \mathbb{P}(\xi^* = k) \end{aligned}$$

where we used that  $\mathbf{u}$  take offsprings according to  $\xi^*$ . Since the ancestors of  $\mathbf{u}$  (between  $\mathbf{u}$  and  $\mathbf{u}'$ ) take offsprings according to  $\hat{\xi}$ , we have

$$\mathbb{P}(d(\mathbf{u}, \mathbf{u}') \text{ is even and } > 0 | \lambda(\mathbf{u}) = \otimes) = \mathbb{P}(\text{Geom}(\eta) \text{ is even and } > 0) = \frac{\eta - 1}{\eta - 2},$$

where

$$\begin{aligned} \eta = \mathbb{P}(\mathbf{G}_{\hat{\xi}} \neq \otimes) &= 1 - \sum_{k \geq 2} \mathbb{P}(\mathbf{G}_k = \otimes) \mathbb{P}(\hat{\xi} = k) \\ &= 1 - \sum_{k \geq 2} \frac{1}{q_k} k q_k t_0^{k-1} = 1 - \frac{t_0(2 - t_0)}{(1 - t_0)^2} = 1 - \kappa(\kappa + 2). \end{aligned}$$

Summing-up

$$\begin{aligned} \mathbb{P}(\lambda(\mathbf{u}) = \otimes, d(\mathbf{u}, \mathbf{u}') \text{ is even}) &= \sum_{k \geq 2} \frac{\eta - 1}{\eta - 2} \mathbb{P}(\mathbf{G}_k = \otimes) \mathbb{P}(\xi^* = k) \\ &= \frac{1}{\sigma^2} \frac{\kappa(\kappa + 2)}{(1 + \kappa)^2} \sum_{k \geq 2} k(k - 1) t_0^{k-1} = \frac{2}{\sigma^2} \kappa^2 (\kappa + 2), \end{aligned} \quad (5.4)$$

where in the last equality we used that  $\sum_{k \geq 2} k(k - 1) t_0^{k-1} = \frac{2t_0}{(1 - t_0)^3}$  and  $t_0 = \frac{\kappa}{1 + \kappa}$ .

Now consider the case where the decoration  $\lambda(\mathbf{u})$  is an  $\mathcal{S}$ -gadget. That is, it consists of a root decorated with a simple permutation with several branches, each of which may be a leaf or a  $\oplus$ -decorated vertex to which at least 2 leaves are attached. By definition, the leaves  $\ell_1$  and  $\ell_2$  are descendants of different children of  $\mathbf{u}$ , say the  $i_1$ -th and  $i_2$ -th. These  $i_1$ -th and  $i_2$ -th branches attached to  $\mathbf{u}$  identify two leaves (the  $i_1$ -th and  $i_2$ -th) of the  $\mathcal{S}$ -gadget  $\lambda(\mathbf{u})$  decorating  $\mathbf{u}$ . If these two leaves belong to the same branch attached to the root of  $\lambda(\mathbf{u})$ , then  $\ell_1$  and  $\ell_2$  do not induce an inversion, as their closest common ancestor in  $\lambda(\mathbf{u})$  is decorated by  $\oplus$  (see also Section 2.5). Otherwise, when they belong to two different branches attached to the root of  $\lambda(\mathbf{u})$  (say the  $j_1$ -th and  $j_2$ -th), it depends on the simple permutation  $\alpha$  appearing in the root of  $\lambda(\mathbf{u})$ :  $\ell_1$  and  $\ell_2$  do not induce an inversion if and only if  $\text{pat}_{\{j_1, j_2\}}(\alpha) = 12$ .

Therefore, in the case where the decoration  $\lambda(\mathbf{u})$  is an  $\mathcal{S}$ -gadget, the limiting probability for  $\ell_1$  and  $\ell_2$  to form a non-inversion is given by

$$\begin{aligned} \mathbb{P}(\lambda(\mathbf{u}) \neq \otimes, i_1, i_2 \text{ are in the same branch of } \lambda(\mathbf{u})) \\ + \mathbb{P}(\lambda(\mathbf{u}) \neq \otimes, i_1, i_2 \text{ are not in the same branch of } \lambda(\mathbf{u}), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12). \end{aligned} \quad (5.5)$$

We start by computing the first probability

$$\begin{aligned} \mathbb{P}(\lambda(\mathbf{u}) \neq \otimes, i_1, i_2 \text{ are in the same branch of } \lambda(\mathbf{u})) \\ = \sum_{k \geq 4} \mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(\mathbf{u}) | \lambda(\mathbf{u}) \neq \otimes, \\ d^+(\mathbf{u}) = k) \mathbb{P}(\lambda(\mathbf{u}) \neq \otimes, d^+(\mathbf{u}) = k) \\ = \sum_{k \geq 4} \mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(\mathbf{u}) | \lambda(\mathbf{u}) \neq \otimes, \\ d^+(\mathbf{u}) = k) \mathbb{P}(\mathbf{G}_k \neq \otimes) \mathbb{P}(\xi^* = k). \end{aligned}$$

Note that, denoting with  $\text{root-deg}(\lambda(u))$  the degree of the root of the decoration of the vertex  $u$ , we have

$$\begin{aligned} & \mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(u) | \lambda(u) \neq \otimes, d^+(u) = k) \\ &= \sum_{a=4}^k \frac{\binom{k-1}{a-1} s_a}{q_k - 1} \mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(u) | \lambda(u) \neq \otimes, \\ & \quad d^+(u) = k, \text{root-deg}(\lambda(u)) = a), \end{aligned}$$

where we use that  $\mathbb{P}(\text{root-deg}(\lambda(u)) = a | \lambda(u) \neq \otimes, d^+(u) = k) = s_a \binom{k-1}{a-1} \frac{1}{q_k - 1}$ , since the number of  $\mathcal{S}$ -gadgets of size  $k$  with root-degree  $a$  is equal to  $s_a$  multiplied by the number of compositions of  $k$  in  $a$  parts. The probability

$$\mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(u) | \lambda(u) \neq \otimes, d^+(u) = k, \text{root-deg}(\lambda(u)) = a)$$

is simply the probability that two uniformly chosen elements a uniform random composition of  $k$  in  $a$  parts (seen as a list of  $k$  elements and  $a - 1$  bars) are in the same part. Summing over the positions of the two uniformly chose, this probability is easily seen to be

$$\binom{k-1}{a-1}^{-1} \binom{k}{2}^{-1} \sum_{i < j} \binom{k - (j - i) - 1}{a - 1} = \frac{2}{a + 1} \cdot \frac{k - a}{k - 1}, \quad (5.6)$$

where the last equality is obtained via a computer algebra system. Summing-up,

$$\begin{aligned} & \mathbb{P}(i_1, i_2 \text{ are in the same branch of } \lambda(u) | \lambda(u) \neq \otimes, d^+(u) = k) \\ &= \frac{1}{q_k - 1} \sum_{a=4}^k s_a \binom{k-1}{a-1} \frac{2}{a + 1} \cdot \frac{k - a}{k - 1} = \frac{2}{q_k - 1} \sum_{a=4}^k \frac{s_a}{a + 1} \binom{k-2}{a-1}, \end{aligned}$$

and so

$$\begin{aligned} & \mathbb{P}(\lambda(u) \neq \otimes, i_1, i_2 \text{ are in the same branch of } \lambda(u)) \\ &= \sum_{k \geq 4} \frac{2}{q_k - 1} \sum_{a=4}^k \frac{s_a}{a + 1} \binom{k-2}{a-1} \mathbb{P}(\mathbf{G}_k \neq \otimes) \mathbb{P}(\xi^* = k) \\ &= \frac{2}{\sigma^2} \sum_{a \geq 4} \frac{s_a}{a + 1} \sum_{k \geq a} k(k-1) \binom{k-2}{a-1} t_0^{k-1} \\ &= \frac{2}{\sigma^2} \frac{t_0}{(1-t_0)^3} \sum_{a \geq 4} s_a a \left(\frac{t_0}{1-t_0}\right)^{a-1} = \frac{2}{\sigma^2} \frac{t_0}{(1-t_0)^3} \mathcal{S}'\left(\frac{t_0}{1-t_0}\right) = \frac{2}{\sigma^2} (\kappa - \kappa^2(\kappa + 2)), \end{aligned} \quad (5.7)$$

where we used the formal power series identity  $\sum_{k \geq a} k(k-1) \binom{k-2}{a-1} t^{k-1} = a(a+1) \frac{t^a}{(1-t)^{a+2}}$  to go from the third to the fourth line, and in the last equality we used Eq. (5.3) and  $\kappa = \frac{t_0}{1-t_0}$ .

It remains to compute the second term in Eq. (5.5). Using the obvious notation  $\mathcal{S}^{\leq k} = \cup_{j \leq k} \mathcal{S}^j$ , we start by determining

$$\begin{aligned} & \mathbb{P}(i_1, i_2 \text{ are not in the same branch of } \lambda(u), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12 | \lambda(u) \neq \otimes, d^+(u) = k) \\ &= \sum_{\alpha \in \mathcal{S}^{\leq k}} \mathbb{P}(i_1, i_2 \text{ are not in the same branch of } \lambda(u), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12 | \lambda(u) \in G_\alpha^k) \\ & \quad \cdot \mathbb{P}(\lambda(u) \in G_\alpha^k | \lambda(u) \neq \otimes, d^+(u) = k), \end{aligned} \quad (5.8)$$

where  $G_\alpha^k$  denotes the set of  $\mathcal{S}$ -gadgets of size  $k$  with root-label  $\alpha$ . Trivially, recalling that  $\binom{k-1}{|\alpha|-1}$  is the number of  $\mathcal{S}$ -gadgets with  $k$  leaves and root decorated by  $\alpha$ , we have

$$\mathbb{P}(\lambda(u) \in G_\alpha^k | \lambda(u) \neq \otimes, d^+(u) = k) = \frac{\binom{k-1}{|\alpha|-1}}{q_k - 1}.$$

Using again the formula (5.6),

$$\begin{aligned} \mathbb{P}(i_1, i_2 \text{ are not in the same branch of } \lambda(u), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12 | \lambda(u) \in G_\alpha^k) \\ = \widetilde{\text{occ}}(12, \alpha) \left( 1 - \frac{2}{|\alpha| + 1} \cdot \frac{k - |\alpha|}{k - 1} \right), \end{aligned}$$

where  $\widetilde{\text{occ}}(12, \alpha) = \binom{|\alpha|}{2}^{-1} \text{occ}(12, \alpha)$  is the probability that two random elements of  $\alpha$  do not form an inversion. Substituting the last two equations in Eq. (5.8) we have

$$\begin{aligned} \mathbb{P}(i_1, i_2 \text{ are not in the same branch of } \lambda(u), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12 | \lambda(u) \neq \otimes, d^+(u) = k) \\ = \sum_{\alpha \in \mathcal{S}^{\leq k}} \widetilde{\text{occ}}(12, \alpha) \left( 1 - \frac{2}{|\alpha| + 1} \cdot \frac{k - |\alpha|}{k - 1} \right) \frac{\binom{k-1}{|\alpha|-1}}{q_k - 1}. \end{aligned}$$

Recalling that  $\mathbb{P}(\lambda(u) \neq \otimes, d^+(u) = k) = \mathbb{P}(\mathbf{G}_k \neq \otimes) \mathbb{P}(\xi^* = k) = (q_k - 1)k(k-1)t_0^{k-1}/\sigma^2$ , it follows that

$$\begin{aligned} \mathbb{P}(\lambda(u) \neq \otimes, i_1, i_2 \text{ are not in the same branch of } \lambda(u), \text{pat}_{\{j_1, j_2\}}(\alpha) = 12) \\ = \frac{1}{\sigma^2} \sum_{k \geq 4} \sum_{\alpha \in \mathcal{S}^{\leq k}} \widetilde{\text{occ}}(12, \alpha) \left( 1 - \frac{2}{|\alpha| + 1} \cdot \frac{k - |\alpha|}{k - 1} \right) \binom{k-1}{|\alpha|-1} k(k-1)t_0^{k-1} \\ = \frac{1}{\sigma^2} \sum_{\alpha \in \mathcal{S}} \widetilde{\text{occ}}(12, \alpha) \sum_{k \geq |\alpha|} \left( \binom{k-1}{|\alpha|-1} - \frac{2}{|\alpha| + 1} \cdot \binom{k-2}{|\alpha|-1} \right) k(k-1)t_0^{k-1} \quad (5.9) \\ = \frac{2}{\sigma^2} \frac{t_0}{(1-t_0)^4} \sum_{\alpha \in \mathcal{S}} \text{occ}(12, \alpha) \left( \frac{t_0}{1-t_0} \right)^{|\alpha|-2} = \frac{2}{\sigma^2} \kappa(1+\kappa)^3 \text{Occ}_{12}(\kappa), \end{aligned}$$

where, to go from the third to the fourth line, we used that  $\text{occ}(12, \alpha) = \widetilde{\text{occ}}(12, \alpha) \binom{|\alpha|}{2}$  and the formal power series identity  $\sum_{k \geq a} \left( \binom{k-1}{a-1} - \frac{2}{a+1} \cdot \binom{k-2}{a-1} \right) k(k-1)t^{k-1} = (a-1)a \frac{t^{a-1}}{(1-t)^{a+2}}$ , and in the last equality we used that  $\kappa = \frac{t_0}{1-t_0}$  (i.e.  $t_0 = \frac{\kappa}{1+\kappa}$ ) and the definition of  $\text{Occ}_{12}$ . Summing up the results in Eqs. (5.4), (5.7) and (5.9) we conclude that Eq. (5.1) holds.  $\square$

## 6 Local convergence

In this section we investigate the local limits of uniform permutations in a fixed substitution-closed class  $\mathcal{C}$ . We work under the following assumption.

**Assumption 6.1.** Consider the associated random variable  $\xi$  defined by Eq. (3.12). We assume that  $\mathbb{E}[\xi] = 1$ .

We highlight that in this section we do not assume the finite variance hypothesis (as done in Section 5). See also Theorem 3.5 for an explicit characterization of this assumption.

Before stating our results, we recall in the following two sections the notions of local convergence for permutations and trees.

## 6.1 Local limits for permutations

In this section we recall the definition of local topology for permutations recently introduced by Borga in [13]. We start by defining finite and infinite *rooted* permutations. Then we introduce a local distance and the corresponding notion of convergence for deterministic sequences of rooted and unrooted permutations. Finally, we extend this notion of convergence (in two non-equivalent ways) to sequences of *random* unrooted permutations.

**Definition 6.2.** A finite rooted permutation is a pair  $(\nu, i)$ , where  $\nu \in \mathfrak{S}^n$  and  $i \in [n]$  for some  $n \in \mathbb{N}$ .

We denote with  $\mathfrak{S}_\bullet^n$  the set of rooted permutations of size  $n$  and with  $\mathfrak{S}_\bullet := \bigcup_{n \in \mathbb{N}} \mathfrak{S}_\bullet^n$  the set of finite rooted permutations. We write sequences of finite rooted permutations in  $\mathfrak{S}_\bullet$  as  $(\nu_n, i_n)_{n \in \mathbb{N}}$ .

To a rooted permutation  $(\nu, i)$ , we associate (as shown in the right-hand side of Fig. 9) the pair  $(A_{\nu, i}, \preccurlyeq_{\nu, i})$ , where  $A_{\nu, i} := [-i + 1, |\nu| - i]$  is a finite interval containing 0 and  $\preccurlyeq_{\nu, i}$  is a total order on  $A_{\nu, i}$ , defined for all  $\ell, j \in A_{\nu, i}$  by

$$\ell \preccurlyeq_{\nu, i} j \quad \text{if and only if} \quad \nu(\ell + i) \leq \nu(j + i).$$

Informally, the elements of  $A_{\nu, i}$  should be thought of as the column indices of the diagram of  $\nu$ , shifted so that the root is in column 0. The order  $\preccurlyeq_{\nu, i}$  then corresponds to the vertical order on the dots in the corresponding columns.

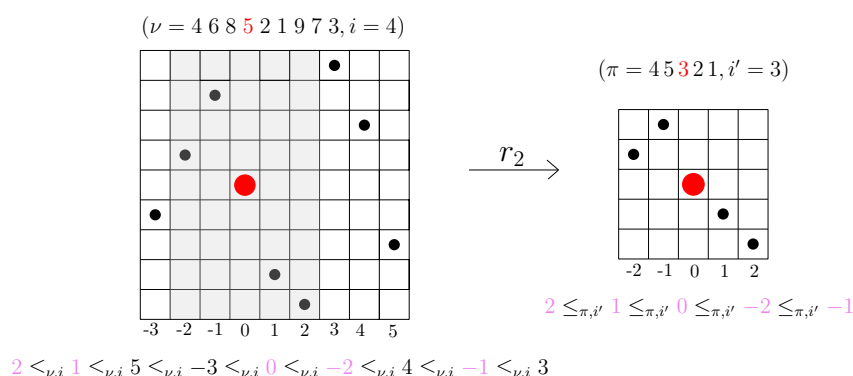


Figure 9: Two rooted permutations and the associated total orders. The big red dot indicates the root of the permutation. The vertical grey strip and the relation between the two rooted permutations will be clarified later.

Clearly this map is a bijection from the space of finite rooted permutations  $\mathfrak{S}_\bullet$  to the space of total orders on finite integer intervals containing zero. Consequently and throughout the paper, we identify every rooted permutation  $(\nu, i)$  with the total order  $(A_{\nu, i}, \preccurlyeq_{\nu, i})$ .

Thanks to the identification between rooted permutations and total orders, the following definition of infinite rooted permutation is natural.

**Definition 6.3.** We call infinite rooted permutation a pair  $(A, \preccurlyeq)$  where  $A$  is an infinite interval of integers containing 0 and  $\preccurlyeq$  is a total order on  $A$ . We denote the set of infinite rooted permutations by  $\mathfrak{S}_\bullet^\infty$ .

We highlight that infinite rooted permutations can be thought of as rooted at 0. We set

$$\tilde{\mathfrak{S}}_\bullet := \mathfrak{S}_\bullet \cup \mathfrak{S}_\bullet^\infty,$$

which is the set of all (finite and infinite) rooted permutations.

We now introduce the following *restriction function around the root* defined, for every  $h \in \mathbb{N}$ , as follows

$$r_h: \tilde{\mathfrak{S}}_{\bullet} \longrightarrow \mathfrak{S}_{\bullet};$$

$$(A, \preccurlyeq) \mapsto (A \cap [-h, h], \preccurlyeq) . \quad (6.1)$$

We can think of restriction functions as a notion of neighbourhood around the root. For finite rooted permutations we also have the equivalent description of the restriction functions  $r_h$  in terms of consecutive patterns: if  $(\nu, i) \in \tilde{\mathfrak{S}}_{\bullet}$  then  $r_h(\nu, i) = (\text{pat}_{[a,b]}(\nu), i)$  where we take  $a = \max\{1, i - h\}$  and  $b = \min\{|\nu|, i + h\}$ .

The *local distance*  $d_p$  on the set of (possibly infinite) rooted permutations  $\tilde{\mathfrak{S}}_{\bullet}$  is defined as follows: given two rooted permutations  $(A_1, \preccurlyeq_1), (A_2, \preccurlyeq_2) \in \tilde{\mathfrak{S}}_{\bullet}$ ,

$$d_p((A_1, \preccurlyeq_1), (A_2, \preccurlyeq_2)) = 2^{-\sup\{h \in \mathbb{N} : r_h(A_1, \preccurlyeq_1) = r_h(A_2, \preccurlyeq_2)\}}, \quad (6.2)$$

with the classical conventions that  $\sup \emptyset = 0$ ,  $\sup \mathbb{N} = +\infty$  and  $2^{-\infty} = 0$ . The metric space  $(\tilde{\mathfrak{S}}_{\bullet}, d_p)$  is a compact space and so Polish, i.e., separable and complete (see [13, Theorem 2.16]).

The above distance entails a notion of convergent sequences of *rooted* permutations. For a sequence  $\nu_n$  of *unrooted* permutations, we consider the sequence of *random* rooted permutations  $(\nu_n, i_n)$ , where  $i_n$  is a uniform random index of  $\nu_n$ . We say that  $\nu_n$  converges in the Benjamini-Schramm sense if the sequence of *random* rooted permutations  $(\nu_n, i_n)$  converges in distribution for the above distance  $d_p$ . This definition is inspired from Benjamini-Schramm convergence for graphs (see [10]).

Benjamini-Schramm convergence can be extended in two different ways for sequences of random permutations  $(\nu_n)_{n \geq 1}$ : the *annealed* and the *quenched* version of the Benjamini-Schramm convergence. These two different versions come from the fact that there are two sources of randomness, one for the choice of the random permutation  $\nu_n$ , and one for the random root  $i_n$ . Intuitively, in the annealed version, the random permutation and the random root are taken simultaneously, while in the quenched version, the random permutation should be thought as frozen when we take the random root.

We now give the formal definitions. In both cases,  $(\nu_n)_{n \in \mathbb{N}}$  denotes a sequence of random permutations in  $\mathfrak{S}$  and  $i_n$  denotes a uniform index of  $\nu_n$ , i.e., a uniform integer in  $[1, |\nu_n|]$ .

**Definition 6.4** (Annealed version of the Benjamini-Schramm convergence). *We say that  $(\nu_n)_{n \in \mathbb{N}}$  converges in the annealed Benjamini-Schramm sense to a random variable  $\nu_{\infty}$  with values in  $\tilde{\mathfrak{S}}_{\bullet}$  if the sequence of random variables  $(\nu_n, i_n)_{n \in \mathbb{N}}$  converges in distribution to  $\nu_{\infty}$  with respect to the local distance  $d_p$ . In this case we write  $\nu_n \xrightarrow{aBS} \nu_{\infty}$  instead of  $(\nu_n, i_n) \xrightarrow{d} \nu_{\infty}$ .*

**Definition 6.5** (Quenched version of the Benjamini-Schramm convergence). *We say that  $(\nu_n)_{n \in \mathbb{N}}$  converges in the quenched Benjamini-Schramm sense to a random measure  $\mu^{\infty}$  on  $\tilde{\mathfrak{S}}_{\bullet}$  if the sequence of conditional laws  $(\mathcal{L}((\nu_n, i_n) | \nu_n))_{n \in \mathbb{N}}$  converges in distribution to  $\mu^{\infty}$  with respect to the weak topology induced by the local distance  $d_p$ . In this case we write  $\nu_n \xrightarrow{qBS} \mu^{\infty}$  instead of  $\mathcal{L}((\nu_n, i_n) | \nu_n) \xrightarrow{d} \mu^{\infty}$ .*

We highlight that, in the annealed version, the limiting object is a *random variable* with values in  $\tilde{\mathfrak{S}}_{\bullet}$ , while for the quenched version, the limiting object  $\mu^{\infty}$  is a *random measure* on  $\tilde{\mathfrak{S}}_{\bullet}$ .

We have the following characterizations of the two versions of the Benjamini-Schramm convergence [13, Section 2.5]. Recall that  $\widetilde{c\text{-occ}}(\pi, \nu)$  denotes the proportion of consecutive occurrences of a pattern  $\pi$  in  $\nu$ , namely,

$$\widetilde{c\text{-occ}}(\pi, \nu) = \frac{c\text{-occ}(\pi, \nu)}{n} = \frac{1}{n} \text{card}\{\text{intervals } I \subseteq [n] \text{ s.t. } \text{pat}_I(\nu) = \pi\}.$$

**Theorem 6.6.** *For any  $n \in \mathbb{N}$ , let  $\nu_n$  be a random permutation of size  $n$ . Then*

- i) *The sequence  $(\nu_n)_{n \in \mathbb{N}}$  converges in the annealed Benjamini–Schramm sense to some  $\nu_\infty$  if and only if there exist non-negative real numbers  $(\Delta_\pi)_{\pi \in \mathfrak{S}}$  such that*

$$\mathbb{E}[\widetilde{c\text{-}occ}(\pi, \nu_n)] \rightarrow \Delta_\pi, \quad \text{for all patterns } \pi \in \mathfrak{S}.$$

- ii) *The sequence  $(\nu_n)_{n \in \mathbb{N}}$  converges in the quenched Benjamini–Schramm sense to some  $\mu^\infty$  if and only if there exist non-negative real random variables  $(\Lambda_\pi)_{\pi \in \mathfrak{S}}$  such that*

$$(\widetilde{c\text{-}occ}(\pi, \nu_n))_{\pi \in \mathfrak{S}} \xrightarrow{d} (\Lambda_\pi)_{\pi \in \mathfrak{S}},$$

*w.r.t. the product topology.*

Since the variables  $\widetilde{c\text{-}occ}(\pi, \nu_n)$  take values in  $[0, 1]$ , the quenched Benjamini–Schramm convergence implies the annealed one.

The goal of the following sections is to prove that a sequence of uniform permutations in a substitution-closed class converges in the quenched Benjamini–Schramm sense using the packed trees representing permutations. To this end, we need to introduce a local topology for trees.

## 6.2 Local limits for decorated trees

In this section we introduce a local topology for decorated trees with a distinguished leaf (called pointed trees in the sequel). This is a straight-forward adaptation of that for trees with a distinguished vertex introduced by Stufler in [51].

Following the presentation in [48, Section 6.3.1], we start by defining an infinite pointed plane tree  $\mathcal{U}_\infty^\bullet$  (see Fig. 10 below). This infinite tree is meant to be a pointed analogue of Ulam–Harris tree, so that pointed trees will be seen as subsets of it. To construct  $\mathcal{U}_\infty^\bullet$ , we take a spine  $(u_i)_{i \geq 0}$  that grows downwards, that is, such that  $u_i$  is the parent of  $u_{i-1}$  for all  $i \geq 1$ . Any vertex  $u_i$ , with  $i \geq 1$ , has an infinite number of children to the left and to the right of its distinguished offspring  $u_{i-1}$ . The former are ordered from right to left and denoted by  $(v_{L,j}^i)_{j \geq 1}$ , the latter are ordered from left to right and denoted by  $(v_{R,j}^i)_{j \geq 1}$ . Each of these vertices not belonging to the spine  $(u_i)_{i \geq 0}$  is the root of a copy of the Ulam–Harris tree  $\mathcal{U}_\infty$ . We always think of  $\mathcal{U}_\infty^\bullet$  as a tree with distinguished leaf  $u_0$ .

**Definition 6.7.** A (possibly infinite) pointed plane tree  $T^\bullet$  is a subset of  $\mathcal{U}_\infty^\bullet$  such that

- $u_0 \in T^\bullet$ .
- if  $u_p \in T^\bullet$  then  $u_i \in T^\bullet$ , for all  $0 \leq i \leq p$ .
- if  $v_{L,q}^i \in T^\bullet$  (resp.  $v_{R,q}^i \in T^\bullet$ ) then  $u^i \in T^\bullet$  and  $v_{L,j}^i \in T^\bullet$  (resp.  $v_{R,j}^i \in T^\bullet$ ) for all  $1 \leq j \leq q$ .
- Any maximal subset of  $T^\bullet$  contained in one of the Ulam–Harris trees  $\mathcal{U}_\infty$  of  $\mathcal{U}_\infty^\bullet$  is a plane tree.

We denote with  $\mathfrak{T}^\bullet$  the space of (possibly infinite) pointed plane trees.

We say that a pointed tree  $T^\bullet$  in  $\mathfrak{T}^\bullet$  is *locally and upwards finite* if every vertex has finite degree and the intersection of  $T^\bullet$  with any one of the Ulam–Harris trees  $\mathcal{U}_\infty$  of  $\mathcal{U}_\infty^\bullet$  is finite. The set of locally and upwards finite pointed trees will be denoted by  $\mathfrak{T}^{\bullet, \text{luf}}$ .

Any finite plane tree  $T$  together with a distinguished leaf  $v_0$  may be interpreted in a canonical way as a pointed plane tree  $T^\bullet$ , such that  $v_0$  is mapped to  $u_0$ . In particular, the



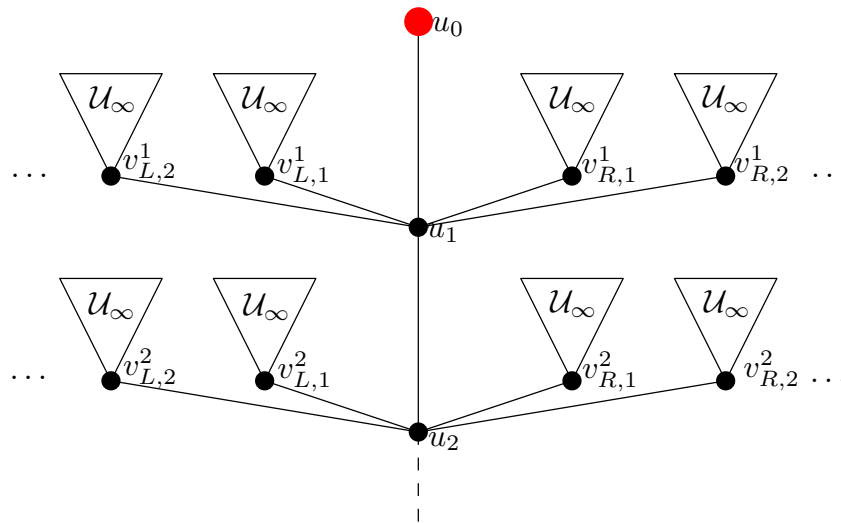


Figure 10: A schema of the infinite plane tree  $\mathcal{U}_\infty^\bullet$ .

backward spine  $u_0, u_1, \dots$  of the associated pointed plane tree  $T^\bullet$  is finite and ends at the root of  $T$ .

Next, we need to extend this notion to decorated trees. Let  $\mathcal{D}$  be a combinatorial class. We define  $\mathcal{D}$ -decorated locally and upwards finite pointed trees, as a tree  $T^\bullet$  in  $\mathfrak{T}^{\bullet, \text{luf}}$ , endowed with a decoration function  $\text{dec} : V_{\text{int}}(T^\bullet) \rightarrow \mathcal{D}$ , such that, for each  $v$  in  $V_{\text{int}}(T^\bullet)$ , the outdegree of  $v$  is exactly  $\text{size}(\text{dec}(v))$ . We denote such a tree with the pair  $(T^\bullet, \text{dec})$  and the space of such trees as  $\mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}}$ . As above, a decorated tree with a distinguished leaf can be identified with an element of this set.

Given a  $\mathcal{D}$ -decorated pointed tree  $(T^\bullet, \lambda_{T^\bullet}) \in \mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}}$ , we denote with  $f_h^\bullet(T^\bullet, \lambda_{T^\bullet})$  the  $\mathcal{D}$ -decorated pointed tree  $(f_h^\bullet(T^\bullet), f_h^\bullet(\lambda_{T^\bullet}))$ , where  $f_h^\bullet(T^\bullet)$  is the pointed fringe subtree rooted at  $u_h$  with distinguished leaf  $u_0$  (if  $u_h$  is not well-defined because  $u_h \notin T^\bullet$ , we set  $f_h^\bullet(T^\bullet) = f_m^\bullet(T^\bullet)$ , where  $m$  is the maximal index such that  $u_m \in T^\bullet$  and  $f_h^\bullet(\lambda_{T^\bullet})$  is  $\lambda_{T^\bullet}$  restricted to the domain  $V_{\text{int}}(f_h^\bullet(T^\bullet))$ . We note that, for any given  $h$ , the image set  $f_h^\bullet(\mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}})$  is countable.

We endow the space  $\mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}}$  with the *local distance*  $d_t$  defined, for all  $(T_1^\bullet, \lambda_{T_1^\bullet}), (T_2^\bullet, \lambda_{T_2^\bullet}) \in \mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}}$ , by

$$d_t((T_1^\bullet, \lambda_{T_1^\bullet}), (T_2^\bullet, \lambda_{T_2^\bullet})) = 2^{-\sup\{h \geq 0 : f_h^\bullet(T_1^\bullet, \lambda_{T_1^\bullet}) = f_h^\bullet(T_2^\bullet, \lambda_{T_2^\bullet})\}}, \quad (6.3)$$

with the classical conventions that  $\sup \emptyset = 0$ ,  $\sup \mathbb{N} = +\infty$  and  $2^{-\infty} = 0$ .

**Remark 6.8.** The distance defined in Eq. (6.3) can be trivially restricted also to the space of non-decorated pointed trees. We point out that this distance is not equivalent to the distance considered in [48, Section 6.3.1] for the space of non-decorated pointed trees. For instance, if  $S_n$ ,  $1 \leq n \leq \infty$  is a star where the root has outdegree  $n$  and its children all have outdegree 0, then the sequence  $(S_n)_{n \geq 1}$  does not converge for our metric (and has no convergent subsequences). This implies that our space is not compact. On the contrary, the space of pointed trees endowed with the distance defined by Stufler in [48] is compact.

We also note (without proof since we do not need this result) that in the subspace of locally finite pointed trees the two distances are topologically equivalent. A proof of this result would be an easy adaptation of [31, Lemma 6.2].

**Proposition 6.9.** *The space  $(\mathfrak{T}_\mathcal{D}^{\bullet, \text{luf}}, d_t)$  is a Polish space.*

*Proof.* The separability is trivial since  $\bigcup_{h \geq 1} f_h^\bullet(\mathfrak{T}_\mathcal{D}^{\bullet, \text{lf}})$  is a countable dense set. The completeness follows from the fact that the space  $\mathfrak{T}_\mathcal{D}^{\bullet, \text{lf}}$  is a closed subspace of a countable product of discrete sets (which is complete) via the map  $(T^\bullet, \lambda_{T^\bullet}) \rightarrow (f_h^\bullet(T^\bullet, \lambda_{T^\bullet}))_{h \geq 1}$ .  $\square$

We end this section defining two versions of the local convergence (similar to those previously defined for permutations) for random decorated trees with a uniform random distinguished leaf. In both definition,  $(T_n, \lambda_n)_{n \in \mathbb{N}}$  is a sequence of random (finite)  $\mathcal{D}$ -decorated trees and  $\ell_n$  is a uniform random leaf of  $(T_n, \lambda_n)$ .

**Definition 6.10** (Annealed Benjamini–Schramm convergence). *We say that  $(T_n, \lambda_n)_{n \in \mathbb{N}}$  converges in the annealed Benjamini–Schramm sense to a random variable  $(T_\infty^\bullet, \lambda_\infty)$  with values in  $\mathfrak{T}_\mathcal{D}^{\bullet, \text{lf}}$  if the sequence of random pointed  $\mathcal{D}$ -decorated trees  $((T_n^\bullet, \lambda_n), \ell_n)_{n \in \mathbb{N}}$  converges in distribution to  $(T_\infty^\bullet, \lambda_\infty)$  with respect to the local distance defined in Eq. (6.3).*

**Definition 6.11** (Quenched Benjamini–Schramm convergence). *We say that  $(T_n, \lambda_n)_{n \in \mathbb{N}}$  converges in the quenched Benjamini–Schramm sense to a random measure  $\mu_\infty^\bullet$  on  $\mathfrak{T}_\mathcal{D}^{\bullet, \text{lf}}$  if the sequence of conditional probability distributions  $\mathcal{L}(((T_n, \lambda_n), \ell_n) | (T_n, \lambda_n))_{n \in \mathbb{N}}$  converges in distribution to  $\mu_\infty^\bullet$  with respect to the weak topology induced by the local distance defined in Eq. (6.3).*

Again, the quenched version is stronger than the annealed one.

**Remark 6.12.** It would also be natural, and closer to the usual notion of Benjamini–Schramm convergence in the literature, to distinguish a uniform random vertex  $v_n$  rather than a uniform random leaf  $\ell_n$  as above. The leaf version is however what we need here for our application to permutations.

### 6.3 Local convergence around a uniform leaf for random packed trees conditioned to the number of leaves

We begin this section by constructing the limiting random pointed packed tree  $P_\infty^\bullet = (T_\infty^\bullet, \lambda_\infty)$ . This tree will be the limit of the sequence of uniform packed trees  $(T_n, \lambda_{T_n})$  considered in Theorem 3.4 pointed at a random leaf.

We recall that  $\xi$  denotes the random variable defined in Eq. (3.12) and  $T$  denotes the associated  $\xi$ -Galton–Watson tree. Additionally, we recall that the random variable  $\hat{\xi}$  defined in Eq. (4.2) is the *size-biased* version of  $\xi$ .

We define the random tree  $T_\infty^\bullet$  in the space  $\mathfrak{T}_\infty^{\bullet, \text{lf}}$  as follows. Let  $u_0$  be the distinguished leaf. For each  $i \geq 1$ , we let  $u_i$  receive offspring according to an independent copy of  $\hat{\xi}$ . The vertex  $u_{i-1}$  gets identified with an offspring of  $u_i$  chosen uniformly at random. All other offspring vertices of  $u_i$  become roots of independent copies of the Galton–Watson tree  $T$ .

Conditionally on  $T_\infty^\bullet$ , the random decoration  $\lambda_\infty(v)$  of each internal vertex  $v$  of  $T_\infty^\bullet$  gets drawn uniformly at random among all  $d_{T_\infty^\bullet}^+(v)$ -sized decorations in  $\widehat{\mathcal{G}(S)}$  independently of all the other decorations ( $\widehat{\mathcal{G}(S)}$  was introduced after Theorem 2.12). This construction yields a random infinite locally and upwards finite pointed packed tree.

We refer to the sequence of (decorated) vertices  $(u_i)_{i \geq 0}$  as the *infinite spine* of  $P_\infty^\bullet = (T_\infty^\bullet, \lambda_\infty)$ .

To simplify notation, we denote the space  $\mathfrak{T}_{\widehat{\mathcal{G}(S)}}^{\bullet, \text{lf}}$  of (possibly infinite) locally and upwards finite pointed packed trees as  $\mathfrak{P}^{\bullet, \text{lf}}$ .

**Proposition 6.13.** *Let  $P_n = (T_n, \lambda_{T_n})$  be the random packed tree considered in Theorem 3.4 and  $P_\infty^\bullet = (T_\infty^\bullet, \lambda_\infty)$  be the limiting random pointed packed tree constructed above. It holds that*

$$\mathcal{L}((P_n, \ell_n) | P_n) \xrightarrow{P} \mathcal{L}(P_\infty^\bullet), \quad (6.4)$$

where  $\ell_n$  is a uniform leaf of  $P_n$  chosen independently of  $P_n$ .

In particular,  $P_n$  converges in the quenched Benjamini–Schramm sense to the deterministic measure  $\mathcal{L}(P_\infty^\bullet)$  and in the annealed Benjamini–Schramm sense to the random tree  $P_\infty^\bullet$ .

Note that the  $\mathcal{L}(P_\infty^\bullet)$  is a measure on  $\mathfrak{P}^{\bullet, \text{anf}}$ . Since the limiting object in quenched Benjamini–Schramm convergence is in general a random measure on  $\mathfrak{P}^{\bullet, \text{anf}}$ , it should be interpreted as a constant random variable, equal to the measure  $\mathcal{L}(P_\infty^\bullet)$ .

*Proof of Theorem 6.13.* The sequence  $\mathcal{L}((P_n, \ell_n)|P_n)_{n \in \mathbb{N}}$  is a sequence of random probability measures on the Polish space  $(\mathfrak{P}^{\bullet, \text{anf}}, d_t)$ . The set of closed and open balls

$$\mathcal{B} = \left\{ B((T^\bullet, \lambda_{T^\bullet}), 2^{-h}) : h \in \mathbb{N}, (T^\bullet, \lambda_{T^\bullet}) \in \mathfrak{P}^{\bullet, \text{anf}} \right\}$$

is a convergence-determining class for the space  $(\mathfrak{P}^{\bullet, \text{anf}}, d_t)$ , i.e., for every probability measure  $\mu$  and every sequence of probability measures  $(\mu_n)_{n \in \mathbb{N}}$  on  $\mathfrak{P}^{\bullet, \text{anf}}$ , the convergence  $\mu_n(B) \rightarrow \mu(B)$  for all  $B \in \mathcal{B}$  implies  $\mu_n \rightarrow \mu$  w.r.t. the weak-topology. This is a trivial consequence of the monotone class theorem and the fact that the intersection of two balls in  $\mathfrak{P}^{\bullet, \text{anf}}$  is either empty or one of them.

Therefore, using [35, Theorem 4.11], the convergence in Eq. (6.4) is equivalent to the following convergence, for all  $k \in \mathbb{N}$  and for all vectors of balls  $(B_i)_{1 \leq i \leq k} \in \mathcal{B}^k$ :

$$\left( \mathcal{L}((P_n, \ell_n)|P_n)(B_i) \right)_{1 \leq i \leq k} \xrightarrow{d} \left( \mathcal{L}(P_\infty^\bullet)(B_i) \right)_{1 \leq i \leq k}.$$

Since the limiting vector in the above equation is deterministic, the above convergence in distribution is equivalent to the convergence in probability. Finally, standard properties of the convergence in probability imply that it is enough to show the component-wise convergence, i.e., for all  $B \in \mathcal{B}$ ,

$$\mathcal{L}((P_n, \ell_n)|P_n)(B) \xrightarrow{P} \mathcal{L}(P_\infty^\bullet)(B). \quad (6.5)$$

Fix a ball  $B = B((T^\bullet, \lambda_{T^\bullet}), 2^{-h}) \in \mathcal{B}$  and note that Eq. (6.5) (which we need to prove) rewrites as

$$\mathbb{P}(f_h^\bullet(P_n, \ell_n) = f_h^\bullet(T^\bullet, \lambda_{T^\bullet}) | P_n) \xrightarrow{P} \mathbb{P}(f_h^\bullet(P_\infty^\bullet) = f_h^\bullet(T^\bullet, \lambda_{T^\bullet})). \quad (6.6)$$

(The left-hand side is a function of  $P_n$ , and hence, a random variable; the right-hand side is a number.)

W.l.o.g. we can assume that  $f_h^\bullet(T^\bullet, \lambda_{T^\bullet}) = (T^\bullet, \lambda_{T^\bullet})$ . Denoting  $\mathcal{L}(P_n)$  the set of leaves of  $P_n$ , the left-hand side writes

$$\begin{aligned} \mathbb{P}(f_h^\bullet(P_n, \ell_n) = (T^\bullet, \lambda_{T^\bullet}) | P_n) &= \frac{|\{\ell \in \mathcal{L}(P_n) : f_h^\bullet(P_n, \ell) = (T^\bullet, \lambda_{T^\bullet})\}|}{n} \\ &= \frac{1}{n} \sum_{\ell \in \mathcal{L}(P_n)} \mathbb{1}_{\{f_h^\bullet(P_n, \ell) = (T^\bullet, \lambda_{T^\bullet})\}} \\ &= \frac{1}{n} \sum_{\ell \in \mathcal{L}(T_n)} \mathbb{1}_{\{f_h^\bullet(T_n, \ell) = T^\bullet\}} \mathbb{1}_{\{\lambda_{f_h^\bullet(T_n, \ell)} = \lambda_{T^\bullet}\}}. \end{aligned} \quad (6.7)$$

For a vertex  $v$  of  $T_n$ , we denote by  $f(T_n, v)$  the fringe subtree rooted at  $v$  and by  $f(\lambda_{(T_n, v)})$  the map  $\lambda|_{V_{\text{int}}(f(T_n, v))}$ . Let also  $T$  be the unpointed version of  $T^\bullet$ . Note that a leaf  $\ell \in \mathcal{L}(T_n)$  satisfies  $f_h^\bullet(T_n, \ell) = T^\bullet$  if and only if its  $h$ -th ancestor  $v$  satisfies  $f(T_n, v) = T$ . Additional, to any  $v$  with  $f(T_n, v) = T$  corresponds exactly one leaf  $\ell$  with  $f_h^\bullet(T_n, \ell) = T^\bullet$  (which

is determined by the pointing). Therefore we can rewrite the last term of the above equation as

$$\frac{1}{n} \sum_{v \in T_n} \mathbb{1}_{\{f(T_n, v)=T\}} \mathbb{1}_{\{f(\lambda_{(T_n, v)})=\lambda_T\}}. \quad (6.8)$$

By [52, Rem. 1.9], we have that

$$\frac{1}{n} \sum_{v \in T_n} \mathbb{1}_{\{f(T_n, v)=T\}} \xrightarrow{P} \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet)).$$

Noting that all fringe subtrees of  $T_n$  that are equal to  $T$  are necessarily disjoint and that, conditioning on  $f(T_n, v) = T$ , then  $f(\lambda_{(T_n, v)}) = \lambda_T$  with probability  $p$ , independently from the rest (specifically  $p = \prod_{u \in T} q_{d_T^+(u)}^{-1}$ ), we can conclude using Chernoff concentration bounds that

$$\begin{aligned} \frac{1}{n} \sum_{v \in T_n} \mathbb{1}_{\{f(T_n, v)=T\}} \mathbb{1}_{\{f(\lambda_{(T_n, v)})=\lambda_T\}} \\ \xrightarrow{P} p \cdot \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet)) = \mathbb{P}(f_h^\bullet(P_\infty^\bullet) = f_h^\bullet(T^\bullet, \lambda_T)), \end{aligned}$$

where the last equality follows from the construction of the map  $\lambda_\infty$ .  $\square$

#### 6.4 The continuity of the bijection between packed trees and $\oplus$ -indecomposable permutations

In this section we consider a substitution-closed class  $\mathcal{C}$  different from the class of separable permutations. The latter case will be considered separately in Section 6.5. We recall that  $\text{DT} := \text{PA} \circ \text{CT}$  is the bijection presented in Theorem 2.18 between  $\oplus$ -indecomposable permutations of  $\mathcal{C}$  and finite packed trees.

The goal of this section is to extend the bijection  $\text{DT}^{-1}$  as a function  $\text{RP}$  from the metric space of (possibly infinite) locally and upwards finite pointed packed trees  $(\mathfrak{P}^{\bullet, \text{inf}}, d_t)$  to the metric space of (possibly infinite) rooted permutations  $(\tilde{\mathfrak{S}}_\bullet, d_p)$ .

First, we need to deal with the introduction of a root in permutations (resp. a pointed leaf in trees) on finite objects. This is very simple, and we extend  $\text{DT}^{-1}$  as a function  $\text{RP}$  from finite pointed packed trees to finite rooted permutations as follows. We recall (see Theorem 2.17) that the  $i$ -th leaf  $\ell$  of a packed tree  $P = \text{DT}(\nu)$  corresponds to the  $i$ -th element of the permutation  $\nu$ . Therefore the following definition is natural:

$$\text{RP}(P, \ell) := (\text{DT}^{-1}(P), i). \quad (6.9)$$

Given an infinite pointed packed tree  $P^\bullet$  with infinitely many  $\mathcal{S}$ -gadget decorations on its infinite spine, we consider the sequence of pointed subtrees

$$(f_{s(h)}^\bullet(P^\bullet))_{h \in \mathbb{N}}$$

consisting of all restrictions for  $s(h) \in \mathbb{N}$  such that  $f_{s(h)}^\bullet(P^\bullet)$  has root decorated with an  $\mathcal{S}$ -gadget.

**Lemma 6.14.** *Let  $P^\bullet$  be an infinite pointed packed tree. Then the (deterministic) sequence of rooted permutations  $(\text{RP}(f_{s(h)}^\bullet(P^\bullet)))_{h \in \mathbb{N}}$  converges in the Benjamini-Schramm sense, as  $h$  tends to  $+\infty$ .*

*Proof.* In Section 2.5, we saw that the pattern associated to a set  $I$  of leaves of a packed tree only depends on any fringe subtree containing all leaves in  $I$  and rooted at a vertex decorated with an  $\mathcal{S}$ -gadget. This implies that the family  $(\text{RP}(f_{s(h)}^\bullet(P^\bullet)))_{h \in \mathbb{N}}$  of elements in  $\mathfrak{S}_\bullet$  is *consistent*, i.e., for all  $h \in \mathbb{N}$ , there exists an integer  $k(h)$  (the half-width of the restriction strip) such that  $r_{k(h)}(\text{RP}(f_{s(h+1)}^\bullet(P^\bullet))) = \text{RP}(f_{s(h)}^\bullet(P^\bullet))$ . By [13, Proposition 2.12], this implies the existence of a limit, which is what we wanted to prove.  $\square$

This lemma allows to define, for an infinite pointed packed tree  $P^\bullet$  having infinitely many  $\mathcal{S}$ -gadget decorations on its infinite spine,

$$\text{RP}(P^\bullet) := \lim_{h \rightarrow \infty} \text{RP}(f_{s(h)}^\bullet(P^\bullet)). \quad (6.10)$$

We now investigate the continuity of the function  $\text{RP}$  with respect to the local topologies. Note that  $\text{RP}$  is defined only for finite pointed packed trees and infinite pointed packed tree with infinitely many  $\mathcal{S}$ -gadget decorations on the infinite spine. This will not be an issue: indeed, we will use the map on a sequence of random pointed packed trees that converges to  $P_\infty^\bullet$  and this limiting pointed packed tree has a.s. infinitely many  $\mathcal{S}$ -gadget decorations on the infinite spine. We start with a definition and a lemma which characterize a certain regularity property of the map  $\text{RP}$ .

**Definition 6.15.** Given a finite plane tree  $T$  we say that a leaf  $\ell_1$  is before (resp. after) a leaf  $\ell_2$  if the post-order label of  $\ell_1$  (see Theorem 2.17) is smaller (resp. greater) than the post-order label of  $\ell_2$ .

**Lemma 6.16.** Fix  $k \geq 0$ . Let  $P_1^\bullet, P_2^\bullet \in \mathfrak{P}^{\bullet, \text{inf}}$  be two pointed packed trees such that for some  $h > 0$  it holds that:

- $f_h^\bullet(P_1^\bullet) = f_h^\bullet(P_2^\bullet)$ ;
- $f_h^\bullet(P_1^\bullet)$  contains at least  $k$  leaves before and  $k$  leaves after the distinguished leaf;
- the root of  $f_h^\bullet(P_1^\bullet)$  is decorated with an  $\mathcal{S}$ -gadget.

Then  $d_p(\text{RP}(P_1^\bullet), \text{RP}(P_2^\bullet)) \leq 2^{-k}$ .

*Proof.* Since  $f_h^\bullet(P_1^\bullet)$  contains at least  $k$  leaves before and  $k$  leaves after the distinguished leaf, and its root is decorated with an  $\mathcal{S}$ -gadget, the restriction  $r_k(\text{RP}(P_1^\bullet))$  is equal to the restriction  $r_k(\text{RP}(f_h^\bullet(P_1^\bullet)))$ . Of course, the same holds for  $P_2^\bullet$ . This follows from the discussion of Section 2.5.

Since  $f_h^\bullet(P_1^\bullet) = f_h^\bullet(P_2^\bullet)$ , it then follows that  $r_k(\text{RP}(P_1^\bullet)) = r_k(\text{RP}(P_2^\bullet))$ . So, by definition of  $d_p$ , we conclude that  $d_p(\text{RP}(P_1^\bullet), \text{RP}(P_2^\bullet)) \leq 2^{-k}$ .  $\square$

We now set

$$C_{\text{RP}} := \{P^\bullet \in \mathfrak{P}^{\bullet, \text{inf}} : \forall k > 0, \exists h(k) > 0 \text{ s.t. } f_{h(k)}^\bullet(P^\bullet) \text{ contains at least } k \text{ leaves before and } k \text{ leaves after the distinguished leaf, and has a root decorated with an } \mathcal{S}\text{-gadget}\}. \quad (6.11)$$

**Proposition 6.17.**  $\text{RP} : (\mathfrak{P}^{\bullet, \text{inf}}, d_t) \rightarrow (\tilde{\mathfrak{S}}_\bullet, d_p)$  is continuous on  $C_{\text{RP}}$ .

*Proof.* Let  $(P_n^\bullet)_{n \geq 0}$  be a convergent sequence in  $(\mathfrak{P}^{\bullet, \text{inf}}, d_t)$  with limit  $P^\bullet \in C_{\text{RP}}$ . Therefore, for all  $k > 0$ ,

- there exist  $N(k) > 0$  such that  $d_t(P_n^\bullet, P^\bullet) \leq 2^{-k}$ , for all  $n \geq N(k)$  (since  $P_n^\bullet \rightarrow P^\bullet$ );
- there exists  $h(k) > 0$  such that  $f_{h(k)}^\bullet(P^\bullet)$  contains at least  $k$  leaves before and  $k$  leaves after the distinguished leaf, and has a root decorated with an  $\mathcal{S}$ -gadget (since  $P^\bullet \in C_{\text{RP}}$ ).

In particular, for all  $k > 0$ , setting  $N'(k) = N(h(k))$  we have that

- $f_{h(k)}^\bullet(P_n^\bullet) = f_{h(k)}^\bullet(P^\bullet)$  for all  $n \geq N'(k)$ ;
- $f_{h(k)}^\bullet(P^\bullet)$  contains at least  $k$  leaves before and  $k$  leaves after the distinguished leaf, and has a root decorated with an  $\mathcal{S}$ -gadget.

Theorem 6.16 implies that, for  $n \geq N'(k)$ , we have  $d_p(\text{RP}(P_n^\bullet), \text{RP}(P^\bullet)) \leq 2^{-k}$ . Since such a  $N'(k)$  exists for all  $k > 0$ , we conclude that  $\text{RP}(P_n^\bullet) \rightarrow \text{RP}(P^\bullet)$ . Therefore the function  $\text{RP}$  is continuous on  $C_{\text{RP}}$ , as claimed.  $\square$

As a final preparation result for the proof of Theorem 1.2 in the non-separable case, we show that the limit object  $P_\infty^\bullet$  is in the continuity set of  $\text{RP}$  with probability 1.

**Proposition 6.18.** *We have  $\mathbb{P}(P_\infty^\bullet \in C_{\text{RP}}) = 1$ .*

*Proof.* Obviously we can rewrite  $\mathbb{P}(P_\infty^\bullet \in C_{\text{RP}})$  as

$$\mathbb{P}(\forall k > 0, \exists h(k) > 0 \text{ s.t. } f_{h(k)}^\bullet(P_\infty^\bullet) \text{ contains at least } k \text{ leaves before and } k \text{ leaves after } u_0, \\ \text{and has a root decorated with an } \mathcal{S}\text{-gadget}).$$

Since the problem is symmetric, it is enough to show that for each fixed  $k > 0$ ,

$$\mathbb{P}(\exists h(k) > 0 \text{ s.t. } f_{h(k)}^\bullet(P_\infty^\bullet) \text{ contains at least } k \text{ leaves before } u_0 \\ \text{and has a root decorated with an } \mathcal{S}\text{-gadget}) = 1.$$

Note that

$$\mathbb{P}(\exists h(k) > 0 \text{ s.t. } f_{h(k)}^\bullet(P_\infty^\bullet) \text{ contains at least } k \text{ leaves before } u_0 \\ \text{and has a root decorated with an } \mathcal{S}\text{-gadget}) \\ \geq \mathbb{P}(P_\infty^\bullet \text{ has at least } k \text{ vertices } u_i \text{ in the infinite spine having at least one left child} \\ \text{and an } \mathcal{S}\text{-gadget as decoration}).$$

Here and after, left child means child to the left of the infinite spine.

By construction, in the infinite tree  $P_\infty^\bullet$ , the vertex  $u_i$  has at least one left child when  $u_{i-1}$  is not identified with its first offspring. Conditioned on  $u_i$  having  $d$  children (which happens with probability  $P(\hat{\xi} = d)$ ), this occurs with probability  $1 - 1/d$ . Moreover, conditioning on  $u_i$  having  $d$  children, the probability that  $u_i$  has an  $\mathcal{S}$ -gadget as decoration is equal to  $\frac{q_d - 1}{q_d}$ , where we recall that  $\mathcal{Q}(z) = \widehat{\mathcal{G}(\mathcal{S})}(z) = \sum_{k \geq 2} q_k z^k$  is the generating in Eq. (3.5), and that  $q_d > 1$  for some  $d$  (since we are not treating the case of separable permutations here). Therefore, for all  $i \geq 1$ ,

$$\mathbb{P}(u_i \text{ has at least one left child} \\ \text{and is decorated by an } \mathcal{S}\text{-gadget}) = \sum_{d \geq 2} \mathbb{P}(\hat{\xi} = d)(1 - 1/d) \frac{q_d - 1}{q_d} > 0.$$

By construction, all these events (for all  $i \geq 1$ ) are independent. Since they happen with some positive probability independent of  $i$ , a.s. at least  $k$  of these events hold. Consequently,  $P_\infty^\bullet$  has a.s. at least  $k$  vertices  $u_i$  in its infinite spine that have at least one left child and are decorated by an  $\mathcal{S}$ -gadget. This concludes the proof.  $\square$

## 6.5 The separable permutations case

For the class of separable permutations, we cannot extend as before the map  $\text{DT}^{-1}$  as a function  $\text{RP}$  from the metric space of (possibly infinite) locally and upwards finite pointed packed trees  $(\mathfrak{P}^{\bullet, \text{luf}}, d_t)$  to the metric space of (possibly infinite) rooted permutations  $(\tilde{\mathfrak{S}}_\bullet, d_p)$ . Indeed, every packed tree obtained from a separable permutation contains only  $\oplus$ -decorations.

Instead, in this case, we have to consider two different functions  $\text{RP}^+$  and  $\text{RP}^-$  from the metric space of (possibly infinite) locally and upwards finite pointed rooted trees

to the metric space of (possibly infinite) rooted permutations. We first define the maps for finite rooted trees pointed at a leaf (where all internal vertices are thought of as decorated by  $\otimes$ ). Let  $(T, \ell)$  be such a tree. We denote with  $(T_{\oplus}, \ell)$  (resp.  $(T_{\ominus}, \ell)$ ) the pointed canonical tree obtained from  $(T, \ell)$  labelling the parent of  $\ell$  with  $\oplus$  (resp.  $\ominus$ ) and then labelling all the other internal vertices in the unique way that prevents the creation of  $\oplus - \oplus$  or  $\ominus - \ominus$  edges. Denoting by  $i$  the label of the leaf  $\ell$  (in the sense of Theorem 2.17), we set

$$\text{RP}^+(T, \ell) := (\text{CT}^{-1}(T_{\oplus}), i) \quad \text{and} \quad \text{RP}^-(T, \ell) := (\text{CT}^{-1}(T_{\ominus}), i). \quad (6.12)$$

Finally, given an infinite pointed tree  $T^\bullet$  we set

$$\text{RP}^+(T^\bullet) := \lim_{h \rightarrow \infty} \text{RP}^+(f_h^\bullet(T^\bullet)) \quad \text{and} \quad \text{RP}^-(T^\bullet) := \lim_{h \rightarrow \infty} \text{RP}^-(f_h^\bullet(T^\bullet)).$$

where the existence of the two limits is justified using similar arguments to the ones used in Theorem 6.14. We now set

$$C_{\text{RP}^*} := \{T^\bullet \in \mathfrak{T}^{\bullet, \text{inf}} : \forall k > 0, \exists h(k) > 0 \text{ s.t. } f_{h(k)}^\bullet(T^\bullet) \text{ contains at least } k \text{ leaves before and } k \text{ leaves after the distinguished leaf.}\}. \quad (6.13)$$

With very similar arguments to the ones used in Theorems 6.17 and 6.18 we have the following.

**Proposition 6.19.** *The functions  $\text{RP}^+ : (\mathfrak{T}^{\bullet, \text{inf}}, d_t) \rightarrow (\tilde{\mathfrak{S}}_\bullet, d_p)$  and  $\text{RP}^- : (\mathfrak{T}^{\bullet, \text{inf}}, d_t) \rightarrow (\tilde{\mathfrak{S}}_\bullet, d_p)$  are continuous on  $C_{\text{RP}^*}$ . Moreover,  $\mathbb{P}(T_\infty^\bullet \in C_{\text{RP}^*}) = 1$ .*

We conclude this section with the following result dealing with the local limit of a uniform canonical tree  $T_n$  associated with separable permutations, conditioned on having  $n$  leaves, and where decorations have been removed. We note that  $T_n$  is distributed as the random packed tree considered in Theorem 3.4 for the case of separable permutations (where decorations, which are all  $\otimes$ , have also been removed). Therefore all the properties for the offspring distribution  $\xi$  are still valid. In particular, we remark that  $\xi$  has finite variance in the case of separable permutations.

**Proposition 6.20.** *Let  $T_n$  be as above and  $T_\infty^\bullet$  be the limiting random pointed tree constructed in Section 6.3. It holds that*

$$\mathcal{L}\left((T_n, \ell_n), (-1)^{\text{ht}(\ell_n)} | T_n\right) \xrightarrow{P} \mathcal{L}(T_\infty^\bullet, B_\pm), \quad (6.14)$$

where  $\ell_n$  is a uniform leaf of  $T_n$  chosen independently of  $T_n$ ,  $\text{ht}(\ell_n)$  denotes the height of the leaf  $\ell_n$  and  $B_\pm$  is a Bernoulli random variable on  $\{1, -1\}$  independent of  $T_\infty^\bullet$ .

In particular,  $T_n$  converges in the quenched Benjamini-Schramm sense to the deterministic measure  $\mathcal{L}(T_\infty^\bullet)$  and in the annealed Benjamini-Schramm sense to the random tree  $T_\infty^\bullet$ .

We highlight that since we want also to keep track of the parity of the distance between the pointed leaf and the root of the tree, Theorem 6.20 does not follow as a simple adaptation from the proof of Theorem 6.13.

*Proof.* With similar arguments to the ones used in the first part of proof of Theorem 6.13, in order to prove Eq. (6.14), it is enough to show that for a fixed leaf-pointed tree  $T^\bullet$ , and for any fixed  $h$ ,

$$\mathbb{P}(f_h^\bullet(T_n, \ell_n) = f_h^\bullet(T^\bullet), (-1)^{\text{ht}(\ell_n)} = 1 | T_n) \xrightarrow{P} \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet), B_\pm = 1). \quad (6.15)$$

Denoting  $\mathfrak{L}(\mathbf{T}_n)$  for the set of leaves of  $\mathbf{T}_n$ , the left-hand side writes

$$\begin{aligned} \mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet), (-1)^{\text{ht}(\ell_n)} = 1 | \mathbf{T}_n) \\ = \frac{|\{\ell \in \mathfrak{L}(\mathbf{T}_n) : f_h^\bullet(\mathbf{T}_n, \ell) = f_h^\bullet(\mathbf{T}^\bullet), (-1)^{\text{ht}(\ell)} = 1\}|}{n} =: N_{\mathbf{T}^\bullet}(n). \end{aligned} \quad (6.16)$$

In order to prove that  $N_{\mathbf{T}^\bullet}(n) \xrightarrow{P} \mathbb{P}(f_h^\bullet(\mathbf{T}_\infty) = f_h^\bullet(\mathbf{T}^\bullet), B_\pm = 1)$ , we use the Second moment method. We start by studying the first moment, which is

$$\mathbb{E}[N_{\mathbf{T}^\bullet}(n)] = \mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet), (-1)^{\text{ht}(\ell_n)} = 1).$$

Using the notation of Section 4.2 and Theorem 4.2, we can rewrite this probability as follows:

$$\begin{aligned} \mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet), (-1)^{\text{ht}(\ell_n)} = 1) \\ = \mathbb{P}(\text{Sh}(c_{\{0\}} \sigma n^{-1/2} . R^{[h]}(\mathbf{T}_n, \ell_n)) \in A_{f_h^\bullet(\mathbf{T}^\bullet)}, (-1)^{\text{ht}(\ell_n)} = 1), \end{aligned}$$

where  $A_{f_h^\bullet(\mathbf{T}^\bullet)}$  is the set of trees  $G$  in  $\mathcal{T}_{1, \{0\}}^{[h]}$  such that at the  $h$ -th ancestor of the distinguished leaf of  $G$  is equal to  $f_h^\bullet(\mathbf{T}^\bullet)$ .

Using Theorem 4.2 with  $\Omega = \{0\}$ ,  $k = 1$ ,  $t = h$  and offspring distribution equal to the one for separable permutations, and the additional result (given by Theorem 4.4) that the parity of the height of  $\ell_n$  converges to a fair coin flip, we have

$$\mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet), (-1)^{\text{ht}(\ell_n)} = 1) \longrightarrow \mathbb{P}(\text{Sh}(\mathbf{T}_{\{0\}}^{1,h}) \in A_{f_h^\bullet(\mathbf{T}^\bullet)}, B_\pm = 1).$$

By comparing the construction of  $\mathbf{T}_{\{0\}}^{1,h}$  in Section 4.3 and that of  $\mathbf{T}_\infty^\bullet$ , we have

$$\mathbb{P}(\text{Sh}(\mathbf{T}_{\{0\}}^{1,h}) \in A_{f_h^\bullet(\mathbf{T}^\bullet)}, B_\pm = 1) = \mathbb{P}(f_h^\bullet(\mathbf{T}_\infty^\bullet) = f_h^\bullet(\mathbf{T}^\bullet), B_\pm = 1).$$

Bringing everything together yields

$$\mathbb{E}[N_{\mathbf{T}^\bullet}(n)] \longrightarrow \mathbb{P}(f_h^\bullet(\mathbf{T}_\infty^\bullet) = f_h^\bullet(\mathbf{T}^\bullet), B_\pm = 1). \quad (6.17)$$

We now study the second moment. We have

$$\mathbb{E}[N_{\mathbf{T}^\bullet}(n)^2] = \mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet) = f_h^\bullet(\mathbf{T}_n, \mathbf{g}_n), (-1)^{\text{ht}(\ell_n)} = 1 = (-1)^{\text{ht}(\mathbf{g}_n)}),$$

where  $\ell_n$  and  $\mathbf{g}_n$  are two uniform random leaves of  $\mathbf{T}_n$ , taking independently conditionally on  $\mathbf{T}_n$ . Again, using the notation of Section 4.2 and Theorem 4.2, we can rewrite this probability as follows:

$$\begin{aligned} \mathbb{P}(f_h^\bullet(\mathbf{T}_n, \ell_n) = f_h^\bullet(\mathbf{T}^\bullet) = f_h^\bullet(\mathbf{T}_n, \mathbf{g}_n), (-1)^{\text{ht}(\ell_n)} = 1 = (-1)^{\text{ht}(\mathbf{g}_n)}) \\ = \mathbb{P}(\text{Sh}(c_{\{0\}} \sigma n^{-1/2} . R^{[h]}(\mathbf{T}_n, (\ell_n, \mathbf{g}_n))) \in B_{f_h^\bullet(\mathbf{T}^\bullet)}, (-1)^{\text{ht}(\ell_n)} = 1 = (-1)^{\text{ht}(\mathbf{g}_n)}), \end{aligned}$$

where  $B_{f_h^\bullet(\mathbf{T}^\bullet)}$  is the set of  $G$  in  $\mathcal{T}_{2, \{0\}}^{[h]}$  such that the two fringe subtrees rooted at the  $h$ -th ancestors of the two distinguished leaves are both equal to  $f_h^\bullet(\mathbf{T}^\bullet)$ .

Using again Theorem 4.2 with  $\Omega = \{0\}$ ,  $t = h$ ,  $k = 2$  and offspring distribution equal to the one for separable permutations, and the additional result (given by Theorem 4.4)



that the parities of the height of  $\ell_n$  and  $g_n$  converges to two independent fair coin flips, we have

$$\begin{aligned} \mathbb{P}(f_h^\bullet(T_n, \ell_n) = f_h^\bullet(T^\bullet) = f_h^\bullet(T_n, g_n), (-1)^{\text{ht}(\ell_n)} = 1 = (-1)^{\text{ht}(g_n)}) \\ \longrightarrow \mathbb{P}(\text{Sh}(T_{\{0\}}^{2,h}) \in B_{f_h^\bullet(T^\bullet)}, B_\pm^1 = 1, B_\pm^2 = 1), \end{aligned}$$

where  $B_\pm^i$ , for  $i = 1, 2$ , are two independent copies of  $B_\pm$ . By construction, in  $T_{\{0\}}^{2,h}$  the neighbourhoods of the two distinguished vertices (here leaves, since  $\Omega = \{0\}$ ) are taken independently so that

$$\mathbb{P}(\text{Sh}(T_{\{0\}}^{2,h}) \in B_{f_h^\bullet(T^\bullet)}, B_\pm^1 = 1, B_\pm^2 = 1) = \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet), B_\pm = 1)^2.$$

Bringing everything together,

$$\mathbb{E}[N_{T^\bullet}(n)^2] \longrightarrow \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet), B_\pm = 1)^2. \quad (6.18)$$

Comparing Eqs. (6.17) and (6.18) and using the standard second moment method, we conclude that

$$N_{T^\bullet}(n) \xrightarrow{P} \mathbb{P}(f_h^\bullet(T_\infty^\bullet) = f_h^\bullet(T^\bullet), B_\pm = 1).$$

Indeed by Chebyshev's inequality, one has, for any fixed  $\varepsilon > 0$ ,

$$\mathbb{P}(|N_{T^\bullet}(n) - \mathbb{E}[N_{T^\bullet}(n)]| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot \text{Var}(N_{T^\bullet}(n)),$$

and the right-hand side tends to zero.  $\square$

## 6.6 Local limit of uniform permutations in substitution-closed classes

We now prove a quenched Benjamini–Schramm convergence result for uniform random permutations in a proper substitution-closed class  $\mathcal{C}$ . As we shall see at the end of the section, this implies our second main result (Theorem 1.2).

**Theorem 6.21.** *Let  $\nu_n$  be a uniform permutation of size  $n$  in a proper substitution-closed class  $\mathcal{C}$ , for all  $n \in \mathbb{N}$ . If  $\mathcal{C}$  is the class of separable permutations, then*

$$\nu_n \xrightarrow{qBS} \mathcal{L}(\text{RP}^{B_\pm}(T_\infty^\bullet)) \quad \text{and} \quad \nu_n \xrightarrow{aBS} \text{RP}^{B_\pm}(T_\infty^\bullet). \quad (6.19)$$

If the set  $S$  of simple permutations in  $\mathcal{C}$  is non-empty and the criticality condition

$$S'(\rho_S) \geq \frac{2}{(1 + \rho_S)^2} - 1 \quad (6.20)$$

is satisfied, then

$$\nu_n \xrightarrow{qBS} \mathcal{L}(\text{RP}(P_\infty^\bullet)) \quad \text{and} \quad \nu_n \xrightarrow{aBS} \text{RP}(P_\infty^\bullet). \quad (6.21)$$

Like after Theorem 6.13, we want to emphasize the nature of the limiting objects above. The limit  $\mathcal{L}(\text{RP}(P_\infty^\bullet))$  (resp. the limit  $\mathcal{L}(\text{RP}^{B_\pm}(T_\infty^\bullet))$ ) is a measure on  $\tilde{\mathfrak{S}}_\bullet$ . Since the limiting object for the quenched Benjamini–Schramm convergence is in general a random measure on  $\tilde{\mathfrak{S}}_\bullet$ , it should be interpreted as a constant random variable, equal to the measure  $\mathcal{L}(\text{RP}(P_\infty^\bullet))$  (resp.  $\mathcal{L}(\text{RP}^{B_\pm}(T_\infty^\bullet))$ ).

*Proof.* We only need to prove the quenched convergence statements, the annealed versions being a simple consequence of the quenched one (see [13, Proposition 2.35]). Moreover, thanks to Theorem 3.2, it is sufficient to prove the statement for a uniform  $\oplus$ -indecomposable permutation  $\nu_n$ .

We first consider the case when  $\mathcal{C}$  is a proper substitution-closed class different from the class of separable permutations. Consider a uniform random leaf  $\ell_n$  in  $P_n$  and a uniform random element  $i_n$  in  $\nu_n$ . We have the following equality in distribution (recall that  $\text{RP}$  denotes the extension of the function  $(\text{PA} \circ \text{CT})^{-1}$  to rooted permutations):

$$(\nu_n, i_n) \stackrel{d}{=} \text{RP}(P_n, \ell_n). \quad (6.22)$$

We analyse the right-hand side conditionally on  $P_n$ . By Theorem 6.13, we know that

$$\mathcal{L}((P_n, \ell_n)|P_n) \xrightarrow{P} \mathcal{L}(P_\infty^\bullet).$$

Moreover, by Theorems 6.17 and 6.18,  $\text{RP}$  is almost surely continuous at  $P_\infty^\bullet$ . Therefore, using a combination of the results stated in [35, Theorem 4.11, Lemma 4.12]<sup>3</sup>,

$$\mathcal{L}(\text{RP}(P_n, \ell_n)|P_n) \xrightarrow{P} \mathcal{L}(\text{RP}(P_\infty^\bullet)).$$

Note that the result described in footnote 3 gives convergence *in distribution*; the limit being a deterministic measure, convergence in probability follows.

Comparing with Eq. (6.22), we have that

$$\mathcal{L}((\nu_n, i_n)|\nu_n) \xrightarrow{P} \mathcal{L}(\text{RP}(P_\infty^\bullet)),$$

which is the quenched convergence in Eq. (6.21).

It remains to prove the theorem for the class of separable permutations. In this case, we have the following equality in distribution (recall that  $\text{RP}^+$  and  $\text{RP}^-$  are the maps defined in Eq. (6.12))

$$(\nu_n, i_n) \stackrel{d}{=} \text{RP}^{\text{sgn}(\ell)}(T_n, \ell_n), \quad (6.23)$$

where  $T_n$  is a uniform undecorated canonical tree with  $n$  leaves,  $\ell_n$  is a uniform leaf of  $T_n$  and  $\text{sgn}(\ell)$  is the sign  $(-1)^{\text{ht}(\ell)}$ .

We analyse the right-hand side conditionally on  $T_n$ . By Theorem 6.20, we know that

$$\mathcal{L}(((T_n, \ell_n), (-1)^{\text{ht}(\ell_n)})|T_n) \xrightarrow{P} \mathcal{L}((T_\infty^\bullet, B_\pm)). \quad (6.24)$$

Moreover, by Theorem 6.19  $\text{RP}^+$  and  $\text{RP}^-$  are almost surely continuous at  $(T_\infty^\bullet, B_\pm)$ . Therefore, using again a combination of the results stated in [35, Theorem 4.11, Lemma 4.12]

$$\mathcal{L}(\text{RP}^{\text{sgn}(\ell_n)}(T_n, \ell_n)|T_n) \xrightarrow{P} \mathcal{L}(\text{RP}^{B_\pm}(T_\infty^\bullet)).$$

Comparing with Eq. (6.23), we have that

$$\mathcal{L}((\nu_n, i_n)|\nu_n) \xrightarrow{P} \mathcal{L}(\text{RP}^{B_\pm}(T_\infty^\bullet)),$$

which is exactly the quenched convergence statement in Eq. (6.19).  $\square$

<sup>3</sup>The specific result that we need is a generalization of the *mapping theorem* for random measures: Let  $(\mu_n)_{n \in \mathbb{N}}$  be a sequence of random measures on a space  $E$  that converges in distribution to a random measure  $\mu$  on  $E$ . Let  $F$  be a function from  $E$  to a second space  $H$  such that the set  $D_F$  of discontinuity points of  $F$  has measure  $\mu(D_F) = 0$  a.s.. Then the sequence of pushforward random measures  $(\mu_n \circ F^{-1})_{n \in \mathbb{N}}$  converges in distribution to the pushforward random measure  $\mu \circ F^{-1}$ .

*Proof of Theorem 1.2.* With the assumption of Theorem 1.2, we just proved (Theorem 6.21) that a uniform permutation  $\nu_n$  in  $\mathcal{C}$  converges in the quenched Benjamini-Schramm sense to some deterministic measure  $\mathcal{L}(\nu_\infty)$ . As recalled in Theorem 6.6 above, the quenched Benjamini-Schramm convergence imply the (joint) convergence of the random variables  $\widetilde{c\text{-}occ}(\pi, \nu_n)$  to some random variables  $\Lambda_\pi$ . Additionally, since the quenched Benjamini-Schramm limit is a deterministic measure, the random variable  $\Lambda_\pi$  are deterministic as well (see [13, Corollary 2.38]), i.e., they are numbers  $\gamma_{\pi, \mathcal{C}}$  in  $[0, 1]$ . This concludes the proof.  $\square$

**Remark 6.22.** Concretely  $\gamma_{\pi, \mathcal{C}}$  is the probability that the restriction of the random order  $\text{RP}(\mathbf{P}_\infty^\bullet)$  (or, in the case of separable permutations,  $\text{RP}^{B\pm}(\mathbf{T}_\infty^\bullet)$ ) on a fixed integer interval of size  $|\pi|$  (e.g.  $[0, |\pi| - 1]$ ) is equal to  $\pi$  (after the identification between permutations and total order on intervals given in Section 6.1). Computing this number involves a sum over countably many configurations of  $\mathbf{P}_\infty^\bullet$  and so it is not immediate, even for simple classes  $\mathcal{C}$  and short patterns  $\pi$ .

## References

- [1] R. Abraham, J.-F. Delmas, and H. Guo, *Critical multi-type Galton–Watson trees conditioned to be large*, J. Theoret. Probab. **31** (2018), no. 2, 757–788. MR-3803914
- [2] M. Albert and M. Atkinson, *Simple permutations and pattern restricted permutations*, Discrete Math. **300** (2005), no. 1, 1–15. MR-2170110
- [3] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab. **1** (1991), no. 2, 228–266. MR-1102319 MR-1102319
- [4] D. Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), no. 1, 1–28. MR-1085326
- [5] D. Aldous, *The continuum random tree. II. An overview*, Stochastic analysis (Durham, 1990), London Math. Soc. Lecture Note Ser., vol. 167, Cambridge Univ. Press, Cambridge, 1991, pp. 23–70. MR-1166406
- [6] D. Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289. MR-1207226
- [7] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, and A. Pierrot, *Universal limits of substitution-closed permutation classes*, Preprint arXiv:1706.08333, to appear in J. Eur. Math. Soc., 2017.
- [8] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, and A. Pierrot, *Scaling limits of permutation classes with a finite specification: a dichotomy*, Preprint arXiv:1903.07522, 2019.
- [9] F. Bassino, M. Bouvel, V. Féray, L. Gerin, and A. Pierrot, *The Brownian limit of separable permutations*, Ann. Probab. **46** (2018), no. 4, 2134–2189. MR-3813988
- [10] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electronic Journal of Probability **6** (2001). MR-1873300
- [11] G. H. Berzunza Ojeda, *On scaling limits of multitype Galton–Watson trees with possibly infinite variance*, ALEA Lat. Am. J. Probab. Math. Stat. **15** (2018), no. 1, 21–48. MR-3748121
- [12] D. I. Bevan, *On the growth of permutation classes*, Ph.D. thesis, The Open University, <https://arxiv.org/abs/1506.06688>, 2015. MR-3347191
- [13] J. Borga, *Local convergence for permutations and local limits for uniform  $\rho$ -avoiding permutations with  $|\rho| = 3$* , Probab. Theory Related Fields **176** (2020), no. 1-2, 449–531. MR-4055194
- [14] J. Borga, E. Duchi, and E. Slivken, *Almost square permutations are typically square*, (2019), Preprint arXiv:1910.04813.
- [15] J. Borga and E. Slivken, *Square permutations are typically rectangular*, Preprint arXiv:1904.03080, to appear in *Annals of Applied Probability*, 2019.
- [16] M. Bouvel, C. Chauve, M. Mishna, and D. Rossin, *Average-case analysis of perfect sorting by reversals*, Discrete Mathematics, Algorithms and Applications **3** (2011), no. 3, 369–392. MR-2855213

- [17] J. Chover, P. Ney, and S. Wainger, *Functions of probability measures*, J. Analyse Math. **26** (1973), 255–302. MR-0348393
- [18] L. de Raphélis, *Scaling limit of multitype Galton–Watson trees with infinitely many types*, Ann. Inst. H. Poincaré Probab. Statist. **53** (2017), no. 1, 200–225. MR-3606739
- [19] L. Devroye and S. Janson, *Distances between pairs of vertices and vertical profile in conditioned Galton–Watson trees*, Random Structures Algorithms **38** (2011), no. 4, 381–395. MR-2829308
- [20] T. Dokos and I. Pak, *The expected shape of random doubly alternating Baxter permutations*, Online J. Anal. Comb. **9** (2014), 12. MR-3238333
- [21] R. Ehrenborg and M. Méndez, *Schröder parenthesizations and chordates*, Journal of Combinatorial Theory, Series A **67** (1994), no. 2, 127–139. MR-1284403
- [22] S. Foss, D. Korshunov, and S. Zachary, *An introduction to heavy-tailed and subexponential distributions*, second ed., Springer Series in Operations Research and Financial Engineering, Springer, New York, 2013. MR-2810144
- [23] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954, Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. MR-0062975
- [24] C. Hoffman, D. Rizzolo, and E. Slivken, *Pattern-avoiding permutations and Brownian excursion Part I: shapes and fluctuations*, Random Structures Algorithms **50** (2017), no. 3, 394–419. MR-3632417
- [25] C. Hoffman, D. Rizzolo, and E. Slivken, *Pattern-avoiding permutations and Brownian excursion, part II: fixed points*, Probab. Theory Related Fields **169** (2017), no. 1-2, 377–424. MR-3704772
- [26] C. Hoffman, D. Rizzolo, and E. Slivken, *Fixed points of 321-avoiding permutations*, Proc. Amer. Math. Soc. **147** (2019), no. 2, 861–872. MR-3894923
- [27] C. Hoffman, D. Rizzolo, and E. Slivken, *Scaling limits of permutations avoiding long decreasing sequences*, Preprint arXiv:1911.04982, 2019.
- [28] C. Holmgren and S. Janson, *Fringe trees, crump–mode–jagers branching processes and  $m$ -ary search trees*, Probab. Surveys **14** (2017), 53–154. MR-3626585
- [29] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. Menezes Sampaio, *Limits of permutation sequences*, J. Combin. Theory Ser. B **103** (2013), no. 1, 93–113. MR-2995721
- [30] S. Janson, *Random cutting and records in deterministic and random trees*, Random Structures Algorithms **29** (2006), no. 2, 139–179. MR-2245498
- [31] S. Janson, *Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation*, Probability Surveys **9** (2012), 103–252. MR-2908619
- [32] S. Janson, *Patterns in random permutations avoiding the pattern 321*, Random Structures & Algorithms **55** (2019), no. 2, 249–270. MR-3983781
- [33] S. Janson, *Patterns in random permutations avoiding some sets of multiple patterns*, Algorithmica **82** (2020), no. 3, 616–641. MR-4058419
- [34] T. Jonsson and S. Stefánsson, *Condensation in nongeneric trees*, J. Stat. Phys. **142** (2011), no. 2, 277–313. MR-2764126
- [35] O. Kallenberg, *Random measures, theory and applications*, Springer, 2017. MR-3642325
- [36] G. Kersting, *On the height profile of a conditioned Galton–Watson tree*, arXiv preprint arXiv:1101.3656 (2011). MR-2908619
- [37] I. Kortchemski, *Invariance principles for Galton–Watson trees conditioned on the number of leaves*, Stochastic Process. Appl. **122** (2012), no. 9, 3126–3172. MR-2946438
- [38] I. Kortchemski, *Limit theorems for conditioned non-generic Galton–Watson trees*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 2, 489–511. MR-3335012
- [39] G. Labelle, *Une nouvelle démonstration combinatoire des formules d’inversion de Lagrange*, Adv. in Math. **42** (1981), no. 3, 217–247. MR-0642392
- [40] M. Maazoun, *On the brownian separable permuton*, Combinatorics, Probability and Computing **29** (2020), no. 2, 241–266. MR-4079636

- [41] A. Marcus and G. Tardos, *Excluded permutation matrices and the Stanley-Wilf conjecture*, J. Combin. Theory Ser. A **107** (2004), no. 1, 153–160. MR-2063960
- [42] G. Miermont, *Invariance principles for spatial multitype Galton–Watson trees*, Ann. Inst. Henri Poincaré Probab. Stat. **44** (2008), no. 6, 1128–1161. MR-2469338
- [43] S. Miner and I. Pak, *The shape of random pattern-avoiding permutations*, Adv. in Appl. Math. **55** (2014), 86–130. MR-3176717
- [44] R. G. Pinsky, *The infinite limit of random permutations avoiding patterns of length three*, Combinatorics, Probability and Computing **29** (2020), no. 1, 137–152. MR-4052932
- [45] J. Pitman and D. Rizzolo, *Schröder’s problems and scaling limits of random trees*, Trans. Amer. Math. Soc. **367** (2015), no. 10, 6943–6969. MR-3378819
- [46] D. Rizzolo, *Scaling limits of Markov branching trees and Galton–Watson trees conditioned on the number of vertices with out-degree in a given set*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 2, 512–532. MR-3335013
- [47] R. Stephenson, *Local convergence of large critical multi-type Galton–Watson trees and applications to random maps*, J. Theoret. Probab. **31** (2018), no. 1, 159–205. MR-3769811
- [48] B. Stufler, *Limits of random tree-like discrete structures*, Preprint arXiv:1612.02580, 2016.
- [49] B. Stufler, *Gibbs partitions: The convergent case*, Random Structures & Algorithms **53** (2018), no. 3, 537–558. MR-3854044
- [50] B. Stufler, *Random enriched trees with applications to random graphs*, Electronic Journal of Combinatorics **25** (2018), no. 3.
- [51] B. Stufler, *Local limits of large Galton–Watson trees rerooted at a random vertex*, Ann. Inst. H. Poincaré Probab. Statist. **55** (2019), no. 1, 155–183. MR-3901644
- [52] B. Stufler, *On the maximal offspring in a subcritical branching process*, Preprint arXiv:1901.04603, 2019.

**Acknowledgments.** JB and MB are partially supported by the Swiss National Science Foundation, under grant number 200021-172536.